Discretized Information and Investor Misperception*

Yongseok Kim† Navid Neshat‡

August 17, 2025

Abstract

In financial markets, information is often presented in simple, categorical formats, such as letter-based credit ratings. This paper shows that such discretized information, even when containing no informational content, can induce investor misperception by creating discontinuities around classification thresholds. Using the South Korean horse racing betting market, we examine horses that are adjacently ranked with nearly identical records within the same race. Unique institutional features of this setting ensure that horses' rankings per se do not affect their performance in subsequent races, a condition rarely met in other contexts. We find that bettors assign significantly higher winning probabilities to horses ranked just above their adjacent peer, a belief contradicted by the two horses' identical subsequent performance. This misperception results in 9.5 percentage points lower returns for bets on higher-ranked horses. Our findings carry important policy implications, indicating that discretized information can mislead investors and distort their decision-making.

*We thank Ehsan Azarmsa, Daniel Carvalho, Nishad Kapadia, Peter Kelly, Kristoph Kleiner, Mathias Kruttli, Alok Kumar, Jordan Martel, Daniel Mochon, Marina Niessner, Chuck Trzcinka, and Mike Woeppel for their helpful comments.

†Freeman School of Business, Tulane University. Email: ykim10@tulane.edu

‡Freeman School of Business, Tulane University. Email: nneshat@tulane.edu

1 Introduction

In financial markets, information is often presented in simplified, discrete formats—such as letter-based credit ratings or mutual fund star ratings. This discretization facilitates decision-making by reducing the cognitive burden associated with processing detailed information (Tversky and Kahneman, 1974; Sims, 2003). However, it can also distort perceptions and lead to errors in judgment when individuals perceive all entities within a category as equivalent—a cognitive shortcut known as categorical thinking (Mullainathan, 2002; Mullainathan *et al.*, 2008; Huang, 2019; Shue *et al.*, 2024).[1] For example, investors may perceive two firms with nearly identical fundamentals as having significantly different credit risk if they fall on opposite sides of a rating threshold. Furthermore, information discretization can induce firms—or issuers more broadly—to strategically adjust investment decisions in order to be placed in a more favorable category (Almeida *et al.*, 2016; Cohen *et al.*, 2024). Such threshold-driven behavior not only distorts corporate policies but also exacerbates capital misallocation by skewing the decisions of categorically minded investors.

Prior research documents that economic decisions often exhibit discontinuities in response to changes in discretized information, such as corporate credit ratings, even when such changes are unrelated to the issuer's underlying fundamentals and carry no intrinsic informational content (e.g., Kliger and Sarig, 2000; Almeida *et al.*, 2017; Cornaggia *et al.*, 2018; Ben-David *et al.*, 2021). Although these discontinuities are often interpreted as evidence of investor misperception, such inferences warrant careful scrutiny. The main empirical challenge is to disentangle perceptual bias from rational responses to the anticipated real effects induced by discretized information (hereafter, the *real effects channel*). Under the real effects channel, for example, even lenders without perceptual biases may respond to a mechanical change in a firm's credit rating for two reasons. First, because credit ratings act as coordination devices for investor beliefs (Boot *et al.*, 2006), a lender may expect others to reduce funding. Second, credit rating changes often carry regulatory implications that deter some lenders from extending credit (Ellul *et al.*, 2011; Chen *et al.*, 2014). Therefore, even when changes in credit ratings are orthogonal to firm fundamentals, lenders may still respond to them due to anticipated changes in the firm's borrowing costs (Manso, 2013; Purnanandam and Wirth, 2023).

[1] For a review of categorical thinking in the psychology literature, see Macrae and Bodenhausen (2000).

This paper contributes to the literature by introducing a novel empirical test that isolates the effect of discretized information on investor misperception from the real effects channel. The key innovation of our approach is to exploit a setting where, unlike most financial markets, changes in discretized information do not alter the actual probability of future outcomes. The South Korean horse racing betting market constitutes such a setting, due to its distinctive institutional characteristics. Crucially, in this setting, the absence of real effects can be empirically validated ex post. Employing this setting, we show that even when changes in discretized information have neither informational value nor real effects, investors still exhibit discontinuities in their economic decisions.

Horse racing betting markets constitute a particularly suitable empirical setting, as they permit researchers to infer individuals' perceived probabilities of future events from observed betting behavior—an advantage seldom available in other contexts.[2] First, the traded assets in these markets are state-contingent claims with a well-defined termination point (Thaler and Ziemba, 1988). This institutional feature allows for the direct elicitation of perceived probabilities from real-stakes decisions and facilitates a comparison between subjective and actual outcome probabilities. In contrast, other empirical designs typically require strong assumptions to infer perceived probabilities or rely on survey-based measures, which lack financial consequences. Second, betting decisions based on perceived probabilities do not influence the realization of outcomes—an uncommon feature in most economic settings (Moskowitz, 2021). For example, in credit markets, a creditor's belief about a firm's probability of default can affect that firm's actual likelihood of default by influencing its access to capital. As a result, even when beliefs are biased, identifying such distortions is complicated by the endogenous relationship between beliefs and outcomes. In horse racing, by contrast, this feedback channel is absent, allowing for cleaner identification of misperception.

Our empirical approach examines whether bettors rely on ranking information even when precise race times are publicly available and easily accessible. Rankings are merely 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 posits that ranking information should not influence the perceptions of rational, unbiased bettors. To

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

implement this approach, we focus on pairs of horses that are adjacently ranked within the same race and exhibit nearly identical race times. We test for discontinuities in betting behavior at ranking thresholds by comparing the wagers placed on the horse ranked just above the threshold to those placed on its immediately lower-ranked counterpart.

However, discontinuities in wagers favoring higher-ranked horses 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. Indeed, this is the case in many horse racing markets around the world, where a *rating system* is used to ensure competitive balance within each race.[3] Within these systems, horses are typically assigned performance-based ratings that are updated after each race. These ratings subsequently 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 investor 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 receive the same 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 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 lack of real effects

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

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 rooted in reliance on redundant categorical information.

Using detailed race-level data from South Korea spanning 1993 to 2024, we construct a dataset at the race×ranking×sign level. For each ranking n within a race, we generate two observations based on relative positioning: one where the nth-ranked horse is designated as the *Higher-Ranked* horse (relative to the horse ranked $n+1$), and another where it is labeled as the *Lower-Ranked* horse (relative to the horse ranked n). Except for horses ranked at the very top or bottom, each horse appears twice—once as higher-ranked and once as lower-ranked—enabling within-race comparisons of adjacent pairs. We restrict attention to pairs of adjacently ranked horses with nearly identical race times—specifically, those separated by a distance of less than or equal to a nose (approximately three inches)—and compare their perceived and actual probabilities of winning in the subsequent race. Perceived probabilities are inferred from betting behavior, under the assumption that bettors place more win bets on horses they believe are more likely to win. Importantly, we do not impose any parametric assumptions on the mapping from bets to beliefs. This model-free approach mitigates concerns related to joint hypothesis testing and avoids erroneous rejection of the null due to model misspecification.

Figure 1 presents the main empirical findings. In each panel, the x-axis plots the difference in race times between horses—that is, higher- and lower-ranked horses around a given ranking with nearly identical race times. 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 suggest that, conditional on race time, ranking provides no incremental information about future performance. However, the betting behavior presented in Panels C and D of Figure 1 reveals a marked divergence from realized performance outcomes. These panels compare the proportions of win and place bets placed on higher- versus lower-ranked horses. Despite exhibiting nearly identical performance ex post, 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 documented discrepancy between perceived and actual probabilities persists even after

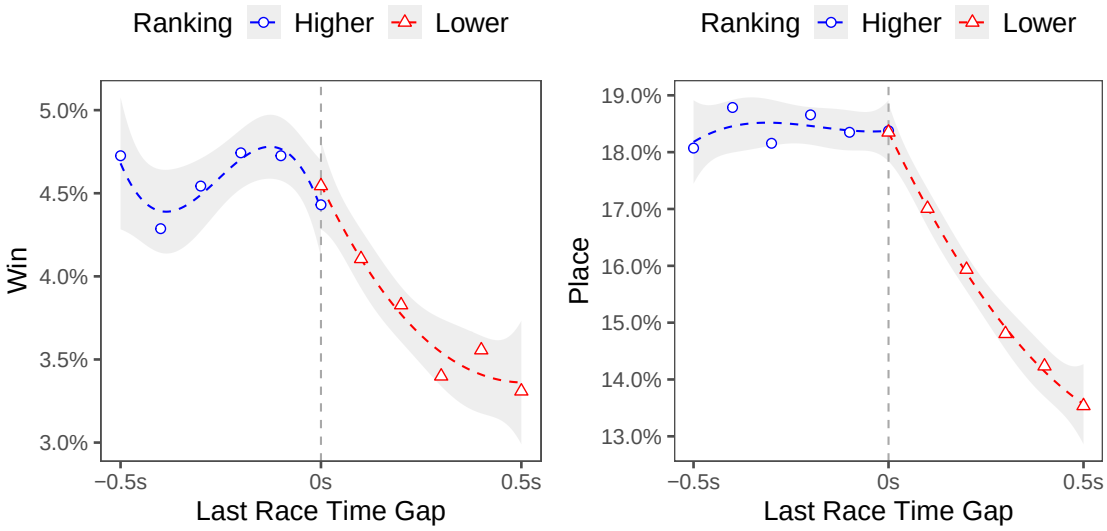
controlling for next-race fixed effects, which account for potential differences in the composition of competing horses. Moreover, there are no observable differences—such as jockey assignment—that would justify the variation in perceived winning probabilities. Alternative explanations based on rational expectations are unlikely to account for both the observed divergence in perceived probabilities and the absence of corresponding differences in actual outcomes.

Our findings indicate that the difference in win-bet shares between higher- and lower-ranked horses can reach up to 0.45 percentage points, equivalent to roughly 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 $0.04 \times (\frac{1}{0.04}) = 1$. However, if bettors overestimate the horse's chances and the betting share increases to 4.4 percent, the expected return falls to $0.04 \times (\frac{1}{0.044}) \approx 0.91$, representing a 9 percentage point decline.

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 examining whether the ranking effect has diminished over time, particularly following technological advancements that have 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.

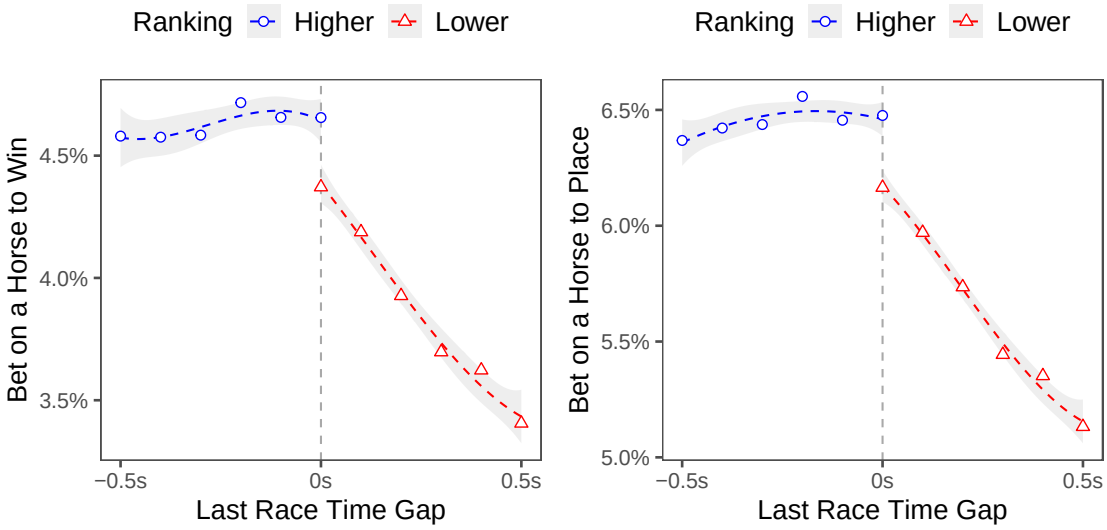
A second potential mechanism is that bettors may rationally disregard race time information due to the cognitive costs associated with processing it—a phenomenon known as rational inattention (Sims, 2003).[4] To test this hypothesis, we examine whether the ranking effect diminishes as economic stakes increase. The rationale is that larger wagers provide greater incentives for bettors to allocate cognitive resources toward processing information, even when doing so is cognitively demanding. Consistent with this prediction, we find that the ranking effect weakens as the average bet size per

[4]For a review of the literature on rational inattention, see Maćkowiak *et al.* (2023).



(A) Probability of Winning

(B) Probability of Placing



(C) Shares of Win Bets

(D) Shares of Place Bets

Figure 1: Ranking, Performance, and Betting Behavior

This figure investigates the extent to which past ranking, conditional on race time, influences subsequent race performance and bettor behavior. The x-axis represents the difference in race times between adjacently ranked horses, calculated as the 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 contrast the shares of win and place bets placed on each horse. Circles and triangles represent unconditional means by last race time gap, while the dashed lines display fitted values from a cubic polynomial. The shaded gray areas denote the standard errors of these fitted lines.

individual increases.

Third, we investigate whether reliance on heuristics contributes to the observed misperception.[5] While the cognitive costs of processing race time information should, in principle, be uniform across ranking thresholds, heuristic reliance may vary depending on 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, like those ranked 12th, which are generally long shots with minimal chances of winning. This asymmetry is consistent with findings in the psychology literature showing that individuals struggle to distinguish between small (e.g., 12th-ranked) and extremely small (e.g., 13th-ranked) probabilities (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996; Snowberg and Wolfers, 2010). Overall, our results support the interpretation that bettor misperception is primarily driven by rational inattention and heuristic decision-making.

Finally, we explore the implications of ranking effects for betting outcomes. While our identification strategy focuses on cases in which horses have virtually identical actual winning probabilities to isolate misperception, reliance on rankings may also influence bettor behavior more broadly. To assess this, we compare the returns of betting strategies involving horses with nearly identical race times but differing rankings. We find that bets placed on higher-ranked horses yield returns up to 9.5 percent lower than those placed on their lower-ranked counterparts. This return differential represents approximately 41 percent of the average loss of 23 cents per dollar wagered. These findings suggest that misperception induced by ranking information can impose substantial economic costs on bettors.

This paper contributes to the literature on betting markets, particularly the strand that examines market efficiency and investor misperceptions (Rosett, 1965; Ali, 1977; Snyder, 1978), in which betting markets have increasingly been used as laboratory settings (e.g., Moskowitz and Vasudevan, 2022; Moskowitz, 2021). A prominent empirical regularity in these markets, first documented by Griffith (1949), is the favorite–longshot bias: bettors systematically overbet longshots and underbet favorites relative to their true winning probabilities, leading to persistent mispricing.[6] Two broad classes of theories have been proposed to explain this bias. Preference-based models

[5] For a review of heuristic decision-making in the psychology literature, see Gigerenzer and Gaissmaier (2011).
[6] For a comprehensive review of the favorite–longshot bias, see Snowberg and Wolfers (2008).

extend standard utility theory by allowing for locally risk-loving behavior (Friedman and Savage, 2024), while perception-based models attribute the bias to cognitive errors (Tversky and Kahneman, 1974; Kahneman and Tversky, 1979). Building on this literature, we leverage distinctive institutional features of the South Korean horse racing market to introduce a new empirical strategy that isolates investor misperception. Specifically, we show that ranking information—discretized yet redundant—significantly influences bettor behavior and induces misperception, even in settings where actual performance outcomes remain constant.

We also extend the literature studying how households form expectations in investment settings. Prior research has documented that household investors frequently rely on discretized information rather than processing the full set of available information (e.g., Choi *et al.*, 2009, 2010; Hartzmark, 2015; Choi and Roberson, 2020; Ben-David *et al.*, 2021). In parallel, recent work highlights the prevalence of gambling preferences among household investors, noting significant overlap between participants in equity and betting markets (e.g., Dorn *et al.*, 2015; Gao and Lin, 2015; Chen *et al.*, 2021; Baker *et al.*, 2024). We advance this literature by showing that household investors' perceptions can be distorted by the provision of discretized information, even when such information is redundant.

Lastly, our findings have important implications for the design of information provision in financial markets. A substantial literature has examined the effects of discretized information on investor behavior, particularly in the contexts of credit ratings and mutual fund ratings.[7] For example, several studies document that investors respond to changes in credit ratings even when those changes are purely mechanical (Kliger and Sarig, 2000; Almeida *et al.*, 2017; Cornaggia *et al.*, 2018). Similarly, mutual funds positioned just above a Morningstar rating threshold receive substantially higher inflows than otherwise similar funds just below the cutoff (e.g., Reuter and Zitzewitz, 2021). Distinct from much of this literature, our paper isolates the effect of discretized information on investor misperception while ruling out explanations based on real effects—that is, cases where investors rationally respond to discontinuities due to their potential economic consequences. Our findings indicate that discretized information, despite being redundant, can

[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)

mislead investors and distort decision-making, highlighting the potential adverse ramifications of information discretization in financial markets.

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 between pairs of horses that differ in ranking but exhibit nearly identical race times. Finally, we highlight the distinct advantages of using betting markets over alternative settings to implement 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:

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

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).

In the context of the favorite-longshot bias, prior research has explored the risk preferences of bettors—for instance, whether they exhibit risk-loving behavior ($u'' > 0$).[8] However, such considerations are not essential for our analysis. We require only that the utility function is strictly increasing—that is, $u' > 0$, so that more is always preferred to less. Under this assumption, if $o_A > o_B$, it follows that $\hat{p}_A < \hat{p}_B$, which enables us to infer bettors' perceived winning probabilities directly from observed betting odds.

[8] For a review of the literature on the favorite–longshot bias, see Thaler and Ziemba (1988) and Snowberg and Wolfers (2008).

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, our primary empirical test is whether:

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

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:

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

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 and thus satisfies $\hat{p}(X) \approx \hat{p}(X+\varepsilon)$ for small $\varepsilon > 0$. Otherwise, discontinuities in perceived probabilities around ranking thresholds could arise even if bettors' perceptions do not depend on N. To address this issue, we restrict our analysis to cases where the actual winning probabilities of horses with nearly identical race times are equal—that is, $p(X) \approx p(X+\varepsilon)$ for small $\varepsilon > 0$. Under this condition, any observed 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 investors' 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 (p). 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 our 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 according to the International Federation of Horseracing Authorities, with 12.7 million betting entrants as reported by the Korea Racing Authority.[9]

Tiering System Many countries with large, developed horse racing markets adopt rating systems to organize races.[10] 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

[9] See International Federation of Horseracing Authorities

[10] 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).

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 competitors. Notably, the primary basis for handicapping was cumulative performance, 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 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 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 help 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.

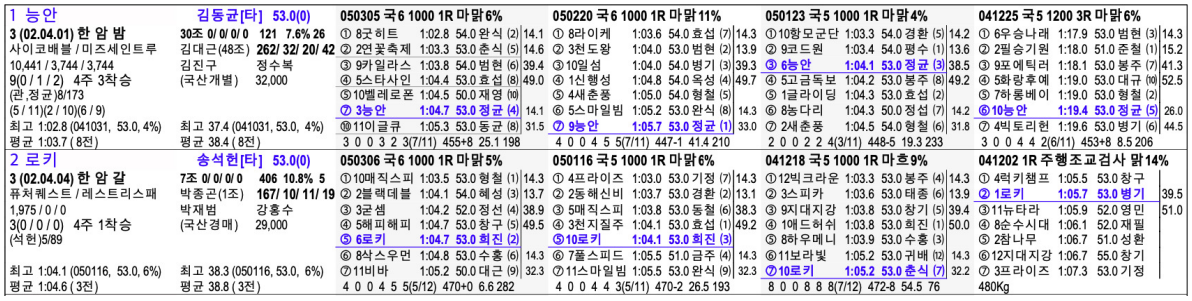
The share of bets on horse i, relative to the total amount of bets, can be derived from the betting odds as follows:

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

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 2 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



1 능안	김동균[타] 53.0(0)	050305 국6 1000 1R 마맑6%					050220 국6 1000 1R 마맑11%					050123 국5 1000 1R 마맑4%					041225 국5 1200 3R 마맑6%				
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 8라이케	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		③ 10일섬	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 마맑5%					050116 국5 1000 1R 마맑6%					041218 국5 1000 1R 마흐9%					041202 1R 주행조교검사 맑14%				
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 3곤샘	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		
(석헌)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 2: 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.

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 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 times measured to the nearest 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

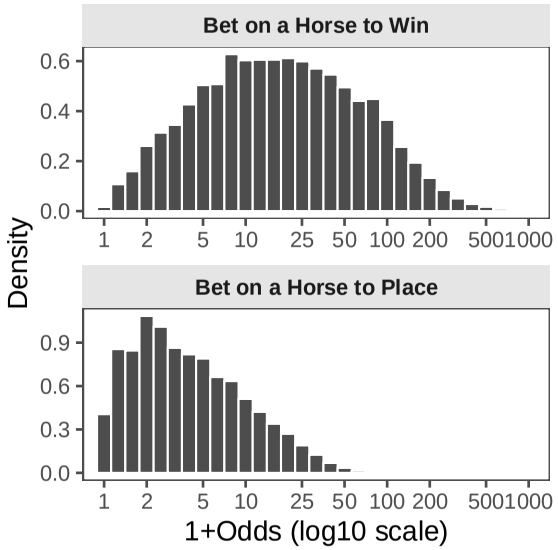
race entries for 46,425 horses across 74,012 races. It contains detailed information for each race, including the horse's weight on the race date, finishing time (measured to the nearest 0.1 second), distance behind the preceding horse, and finalized betting odds.

Figure 3 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 long shots 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).

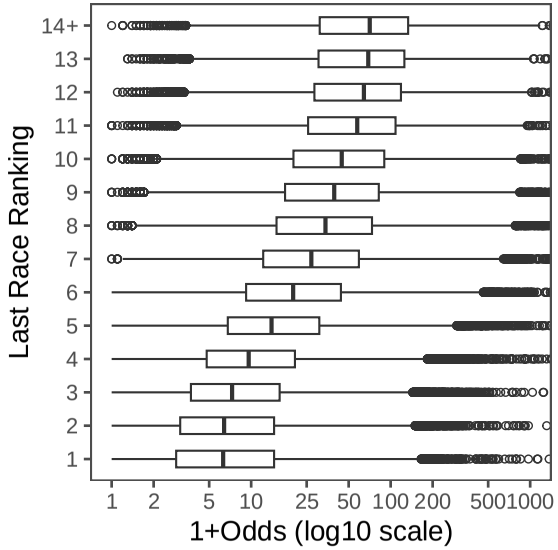
We also collect data on the size distribution of bets from the Public Data Portal (`http://data.go.kr`), maintained by the South Korean government. This dataset spans the period from January 1995 to May 2024 and reports the amount of bets placed at a racecourse on a given race date, categorized into 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 KRW, which is the maximum allowable bet. For each interval, we take the midpoint and compute the average bet size for each racecourse-race date, weighted by the total amount wagered in each interval. This weighted average serves as a proxy for bettors' economic stakes in predicting winning and placing outcomes.

Sample Construction Using the raw datasets described above, we construct a dataset at the race×ranking×sign level. In this dataset, each ranking within a given race generates two observations, one for each sign. Specifically, a horse ranked n in a given race generates two observations, corresponding to each sign. Specifically, a horse ranked n in race i appears twice: once as the higher-ranked horse in the pair $(n, n+1)$, and once as the lower-ranked horse in the pair $(n-1, n)$. Consequently, all horses—except those ranked first or last—are represented twice in the dataset.

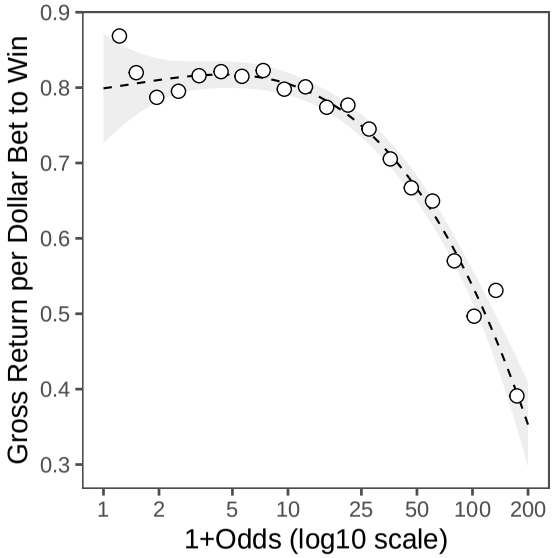
For each race×ranking, we refer to the two adjacent horses around the ranking as a *pair* and



(A) Density Plot of Win and Place Betting Odds



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



(C) Long-Shot Bias

Figure 3: 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.

analyze differences in their subsequent race performance. We exclude observations with missing win betting odds or cases in which horses were scratched or did not finish their subsequent race. This reduces 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. **Race Time Gap**: We begin by selecting pairs whose race time gap in their most recent race is 0.5 seconds or less. This ensures that the two horses had similar performance levels around the ranking threshold. This filter reduces the sample to 919,482 observations.

2. **Rating System**: Next, 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 685,530 observations. Additionally, before February 2015, some races featured handicap weights assigned at the handicappers' discretion. Excluding these races further reduces the sample to 568,189 observations.

3. **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 253,423 observations.

4. **Special Cases and Outliers**: Finally, we remove three types of cases: (i) horses that ran in a "special" race between their previous and next race; (ii) horses whose racing tier changed without explanation between two races; and (iii) horses whose time gap between races falls below the 1st percentile or above the 99th percentile of observed intervals. These exclusions eliminate outliers and special cases, resulting in a final sample of 228,135 observations.

We refer to this refined dataset as the *regression sample* and use it for our empirical analysis. Figure 4 reports the number of observations in both the full sample and the regression sample,

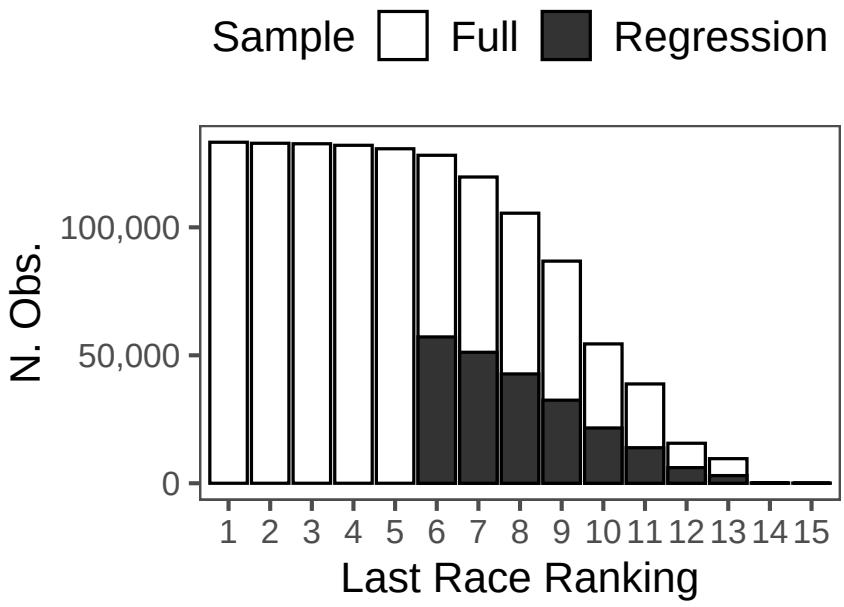


Figure 4: Sample Size by Prior-Race Ranking

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

stratified by ranking in the prior race. In addition, the summary statistics for the regression sample are presented in Table 2.

4 Empirical Model

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

$$Y_{h,i,r,n} = \alpha + \tau D_{h,r,n} + f(\tilde{X}_{h,r,n}) + \mu_{r,n} + \varepsilon_{h,i,r,n}, \tag{6}$$

where $Y_{h,i,r,n}$ denotes the outcome variable for horse h in race i, whose previous race was race r, where it was ranked n. Let $X_{r,n}$ represent the race time of the nth-ranked horse in race r, and $\tilde{X}_{h,r,n}$ denote the gap between horse h's race time and the ranking threshold around position n in race r. When horse h was the higher-ranked of the two around threshold n (i.e., ranked n), then

$\tilde{X}_{h,r,n} = X_{r,n} - X_{r,n+1}$ yields a negative value. Conversely, if horse h was the lower-ranked of the two around ranking n (i.e., ranked $n+1$), then $\tilde{X}_{h,r,n} = X_{r,n+1} - X_{r,n}$ takes a positive value.

The variable $D_{h,r,n}$ is an indicator equal to one if horse h is the higher-ranked horse around ranking n in race r, i.e., $1(\tilde{X}_{h,r,n} < 0)$. The function $f(\tilde{X}_{h,r,n})$ is a flexible function of $\tilde{X}_{h,r,n}$ to control for the running variable's effects. 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 allow for more refined comparisons between the two horses adjacent to a given ranking within the same race.

We employ both non-parametric and parametric approaches to control for the gap between a horse's race time and the race time threshold around its ranking, $\tilde{X}_{h,r,n}$. In the non-parametric approach, we restrict the sample to horse pairs with a distance gap of less than or equal to a nose (approximately 3 inches) in their previous race. These pairs exhibit a race time gap of less than 0.1 seconds, which is recorded as 0 in the dataset. In the parametric approach, we model $f(\tilde{X}_{h,r,n})$ as a polynomial function of $\tilde{X}_{h,r,n}$, interacted with the treatment indicator $D_{r,n}$, as follows:

$$\begin{aligned} f(\tilde{X}_{h,r,n}) = \gamma_1 \tilde{X}_{r,n} + \gamma_2 \tilde{X}_{r,n}^2 + \gamma_3 \tilde{X}_{r,n}^3 \\ + \gamma_4 \left(\tilde{X}_{r,n} \times D_{r,n} \right) + \gamma_5 \left(\tilde{X}_{r,n}^2 \times D_{r,n} \right) + \gamma_6 \left(\tilde{X}_{r,n}^3 \times D_{r,n} \right). \end{aligned} \tag{7}$$

This parametric approach offers flexibility by controlling for potential nonlinearities and interaction effects with the higher-ranked indicator.

One remaining concern is that selection into the next race may introduce 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 issue, we follow Abowd *et al.* (1999) and estimate each jockey's ability using the raw race record dataset, as follows:

$$Y_{h,j,r} = \alpha_h + \lambda_j + \delta_r + \varepsilon_{h,j,r}, \tag{8}$$

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 isolates 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 next race.

Second, it remains possible that higher- and lower-ranked horses are placed in systematically different competitive environments. Although the construction of the regression sample is intended to mitigate this concern, we further address it by including next-race fixed effects, ρ_i, in our robustness checks. These fixed effects account for any potential differences in the competitive environment between the two horses in their next race.

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 5 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 6.5 out of 100 times in their next start, while those ranked 13th win only about 1.9 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

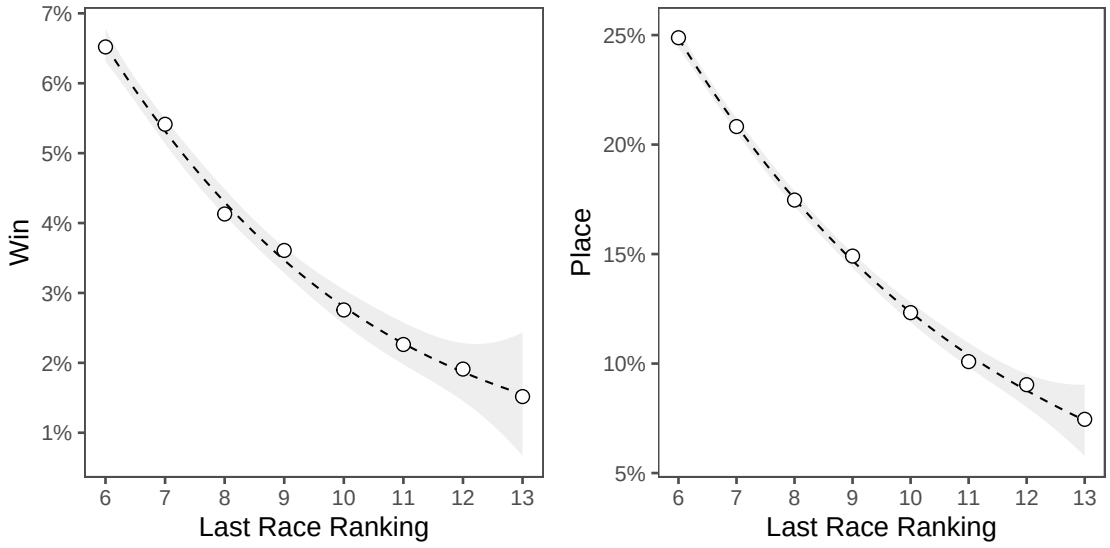
of their jockeys in the subsequent race. To do so, we estimate Equation (6), using either handicap weights or estimated jockey ability as the dependent variable. Table 4 reports the results in two panels. In both panels, columns (1) and (2) focus on the subsample of pairs with a distance gap of less than or equal to a nose. Columns (3) and (4) extend the analysis to pairs with a gap of 0.5 seconds or less, 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 for the higher-ranked horse, as specified in Equation (7). Columns (2) and (4) also include pair fixed effects.

Panel A of Table 4 presents the results for handicap weights. Across all four specifications, the coefficient estimates are not statistically distinguishable from zero, indicating that handicappers do not assign different weights to higher- and lower-ranked horses in their next race. This supports the idea that being marginally higher or lower than a ranking threshold does not influence handicap decisions.

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 (8). 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 (6) 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 time gaps of 0.5 seconds or less, incorporating a polynomial function of the gap between a horse's race time and the race time threshold around its ranking interacted with the higher-ranked indicator, as outlined in Equation (7).

The results are reported in Table 5. Panel A examines winning probabilities, and Panel B



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

Figure 5: 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.

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 (6) 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 gap of 0.5 seconds or less, 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 for the higher-ranked horse, as specified in Equation (7). Columns (2) and (4) also include pair fixed effects.

The results are presented in Table 6. 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 0.45 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 $0.04 \times (\frac{1}{0.04}) = 1$. 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 $0.04 \times (\frac{1}{0.044}) \approx 0.91$, representing a 9 percent loss relative to the scenario without misperception.

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 0.39 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 6.1 percent.

Overall, the results in Table 6 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 5.2, 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 Selections in Competitive Environments

While we carefully construct the regression sample to compare higher- and lower-ranked horses with identical ex ante probabilities of winning and placing in their next race, one may still be concerned about residual selection into competitive environments. For instance, if owners of higher-ranked horses are systematically better at entering races with weaker competition, bettors may appear to respond to rankings when, in fact, they are reacting *rationally* to underlying differences in race competitive environment. Although the absence of differences in actual winning and placing probabilities suggests that this is unlikely, we further address this concern by including next-race fixed effects in the regression specification shown in Equation (6). These fixed effects absorb any potential differences in the competitive environment that higher- and lower-ranked horses face in their next race.

The results are presented in Table 7. Columns (1) and (2) report results for actual winning and placing probabilities, while Columns (3) and (4) examine perceived probabilities based on betting behavior. To ensure a sufficient number of horses in each subsequent race, we rely exclusively on the parametric approach specified in Equation (7) to control for the gap between a horse's race time and the race time threshold around its ranking. All four regression specifications include pair fixed effects.

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 potential differences in the competitive environments faced by higher- and lower-ranked horses in their next race are unlikely to explain the observed patterns.

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

$$Y_{h,i,r,n} = \alpha + \sum_{k \in \{1990s, 2000s, 2010s, 2020\}} \tau_k \left(D_{h,r,n} \times 1\{i \in k\} \right) + f\left(\tilde{X}_{h,r,n}\right) + \mu_{r,n} + \varepsilon_{h,i,r,n}, \tag{9}$$

where $1\{i \in k\}$ is an indicator variable equal to one if the year of race i falls within decade k. All other covariates are defined as in Equation (6).

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 gap of 0.5 seconds or less and 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, as specified in Equation (7).

In all specifications, we include pair fixed effects to absorb time-invariant heterogeneity across pairs.

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, medium, and high bet sizes using the following regression specification:

$$\begin{aligned} Y_{h,i,r,n} = \alpha &+ \tau_L \left(D_{h,r,n} \times \text{Low Bet Size}\right) + \tau_M \left(D_{h,r,n} \times \text{Middle Bet Size}\right) \\ &+ \tau_H \left(D_{h,r,n} \times \text{High Bet Size}\right) + f(\tilde{X}_{h,r,n}) + \mu_{r,n} + \sum_{k \in \{\text{Low,Middle,High}\}} \rho_k + \varepsilon_{h,i,r,n}, \end{aligned} \tag{10}$$

where Low Bet Size is an indicator for whether the average bet size on the race date of race i falls within the bottom 30 percent of the distribution for that racecourse in the given year. Middle Bet Size corresponds to the middle 40 percent, and High Bet Size to the top 30 percent. We also include fixed effects for these bet size categories, denoted by ρ_k. All other covariates are defined as in Equation (6).

The results are presented in Table 9. Mirroring the structure of Table 8, the dependent 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 gap of 0.5 seconds or less and control for the running variable by fitting a polynomial function, as specified in Equation (7). All regression specifications include pair fixed effects.

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. 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. Taken together, the evidence in Table 9 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 long shots, 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 1, 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.

Next, we formally test this pattern by re-estimating Equation (6), interacting the parameter of interest, τ, with indicators for each ranking position. For example, the variable Higher-Ranked (9th) is an indicator equal to one when the higher-ranked horse in the pair is ranked ninth in its race. The results are presented in Table 10.

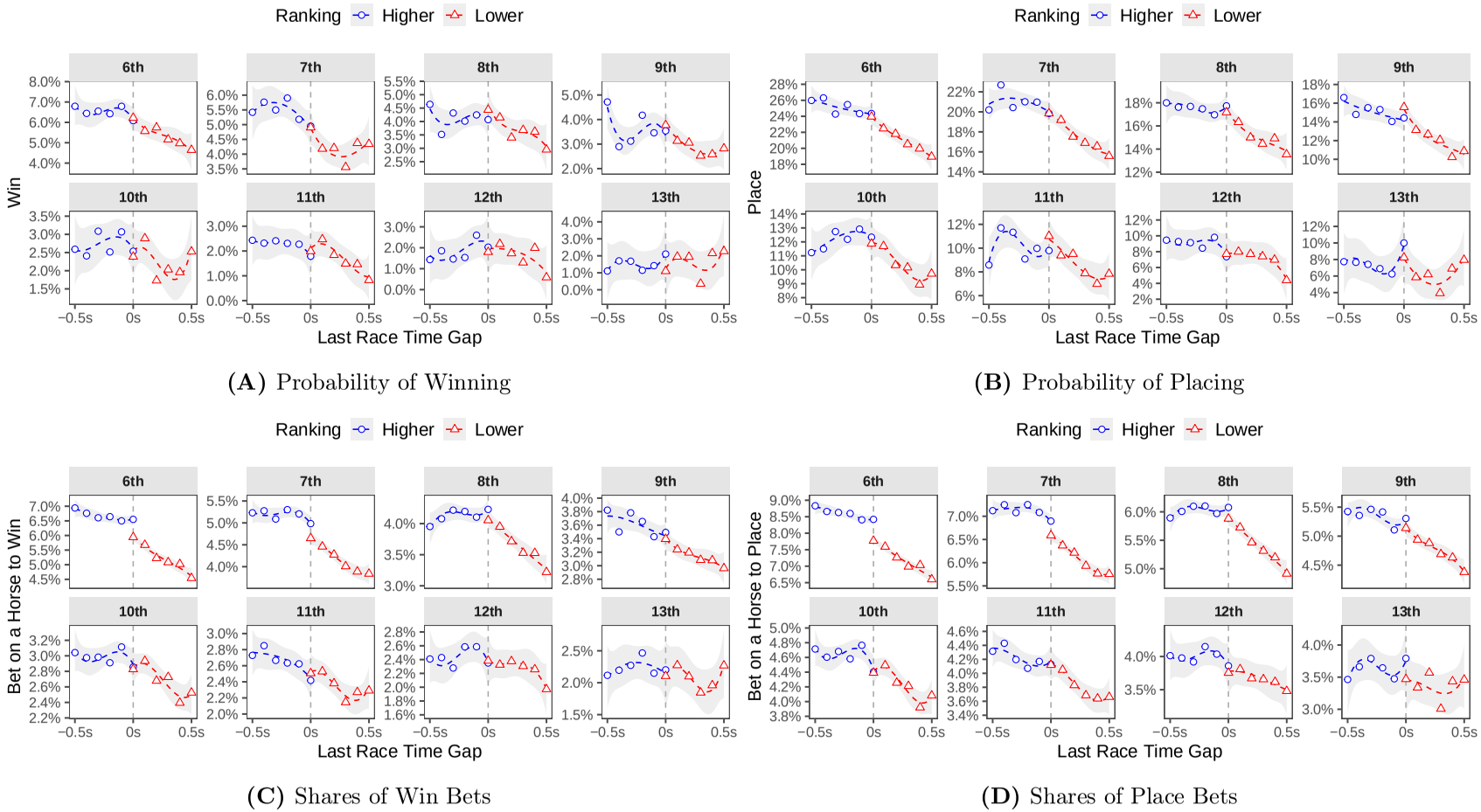
Following the structure of Table 9, the dependent variable in Columns (1) and (2) is the perceived probability of winning, while in Columns (3) and (4), it is the perceived probability of placing. To account for the effects of the running variable, we use both non-parametric and parametric approaches. Columns (1) and (3) restrict the sample to horse pairs with a distance gap of less than or equal to a nose. Columns (2) and (4) expand the sample to include pairs with a gap of up to 0.5 seconds and control for the running variable by fitting a polynomial function, as specified in Equation (7). All specifications include pair fixed effects to absorb unobserved heterogeneity across pairs.

The findings confirm the visual patterns observed in Figure 6. The ranking effects are concentrated at higher ranking positions: discontinuities in perceived winning and placing probabilities are most pronounced among horses near the top of the unawarded 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 in Table 10 suggests that heuristic decision-making can, at least in part, account for the observed misperception in bettor behavior.

5.5 Implications for Betting Strategies

In the preceding sections, our analysis has focused on the regression sample, which comprises adjacently ranked horses with virtually identical expected performance in their subsequent race. This design allows for clean identification of the causal effect of ranking information on bettor perceptions. However, while this sample is well-suited for identification, the implications of such misperceptions likely extend beyond these specific cases. In particular, one would expect these perceptual biases to influence betting strategies across a broader set of ranking positions.

In this section, we turn to the full sample and analyze realized returns from betting on higher- and



(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 1, disaggregated by ranking position. The x-axis in all panels represents the race time difference between adjacently ranked horses. 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. Circles and triangles represent unconditional means of outcome variables by race time gap in the last race, while the dashed lines show fitted values from a cubic polynomial. The gray shaded areas denote the 95 percent confidence intervals for the fitted lines.

lower-ranked horses. The results are presented in Figure 7. Panels A and B plot the average gross return per dollar wagered on win and place bets, respectively, by a horse's ranking in its previous race. The patterns indicate that bets on higher-ranked horses generally yield higher returns than those on lower-ranked horses—a result consistent with the presence of long shot bias in this market.

However, Panels C and D reveal a striking contrast. These panels plot the average gross return per dollar wagered on win and place bets, respectively, for higher- and lower-ranked horses with nearly identical race times in their previous race. Once we condition on race time, the pattern reverses: bets on higher-ranked horses yield significantly lower returns compared to those on lower-ranked horses, despite similar underlying performance.

We formally test this pattern using a regression discontinuity (RD) design. The results are presented in Table 11. The dependent variable in Columns (1) and (2) is Payoff per Dollar Bet to Win, defined as the realized payoff per dollar wagered on a horse to win. Columns (3) and (4) report results for Payoff per Dollar Bet to Place, the realized return per dollar wagered on a horse to place. As in prior analyses, we account for the running variable using both non-parametric and parametric approaches. Columns (1) and (3) restrict the sample to horse pairs with a distance gap of less than or equal to a nose, while Columns (2) and (4) expand the sample to include pairs with a gap of up to 0.5 seconds, controlling for the running variable using a polynomial function as specified in Equation (7). All specifications include pair fixed effects to absorb unobserved heterogeneity across pairs.

To mitigate the influence of extreme values, we estimate the RD models across three ranges of betting odds. Panel A restricts the sample to horses with odds below 50, Panel B extends the threshold to odds below 100, and Panel C further includes horses with odds below 200. Across all panels, the results consistently show that betting on higher-ranked horses yields significantly lower returns. For instance, in Panel A, Column (1), betting on higher-ranked horses to win reduces realized payoff by 9.5 percentage points—equivalent to a 41 percent increase in the average loss of 23 cents per dollar wagered. Similarly, Column (3) shows a 3.8 percentage points reduction in payoff for place bets on higher-ranked horses.

Overall, the results in Table 11 demonstrate that reliance on ranking information imposes a substantial cost on bettors, reducing returns on both win and place bets. These losses are especially pronounced when such reliance leads to discontinuous perceptions of performance around ranking

thresholds—despite the fact that horses straddling these thresholds exhibit nearly identical race times. These findings underscore the economic costs associated with perceptual biases driven by reliance on simple yet redundant signals such as ranking information.

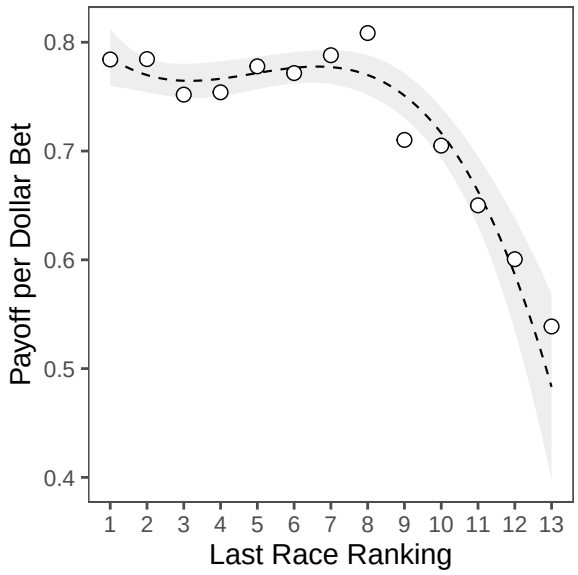
6 Conclusion

The provision of information in simplified, discrete formats is common practice in financial markets. While such discretization can facilitate decision-making by reducing cognitive load, it also carries the risk of distorting investor perceptions and leading to suboptimal outcomes. Leveraging the unique institutional features of the South Korean horse racing market, this paper demonstrates how discretized information—specifically rankings—can misguide investors. By focusing on horses that straddle ranking thresholds but exhibit virtually identical race times, we isolate the effect of ranking information and show that bettors systematically overestimate the chances of horses just above the threshold, despite the absence of any substantive difference in actual performance. This misperception has significant economic consequences, increasing the average bettor's loss by as much as 41 percent.

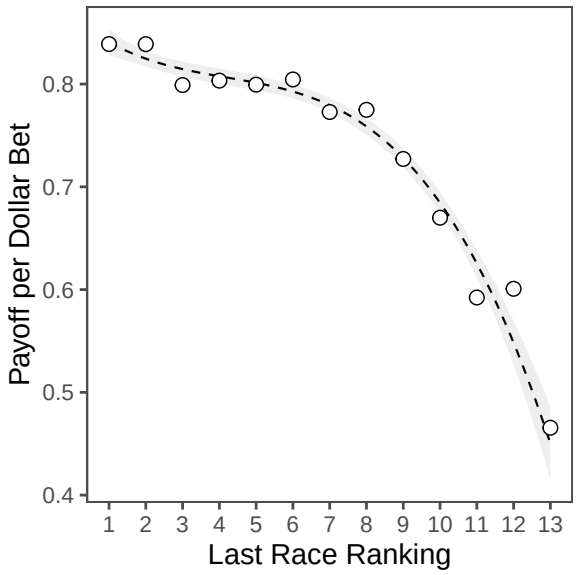
A key contribution of this study is its ability to isolate perceptual bias arising from discretized information. While prior research has shown that market participants respond to mechanical shifts in categorical labels, it remains unclear to what extent such responses simply reflect rational expectations. In the case of corporate credit ratings, for example, a negative change in a firm's rating—even if uncorrelated with fundamentals—could nonetheless increase bankruptcy risk because ratings function as a coordination mechanism among investors. We overcome this empirical challenge by leveraging a unique setting where changes in discretized information *per se* do not affect the actual probability of future outcomes.

Finally, we shed light on the mechanisms underlying this misperception. The perceptual bias induced by discretized information persists even in settings where information acquisition costs are low, suggesting that such costs are unlikely to be the primary driver of our findings. Instead, the evidence is more consistent with explanations based on rational inattention and heuristic decision-making, whereby individuals use discretized information to guide their decisions.

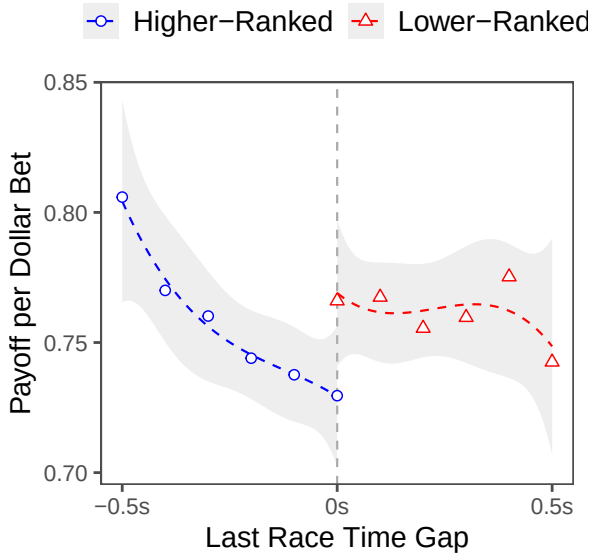
Our findings have important implications for financial markets. Commonly used discretized



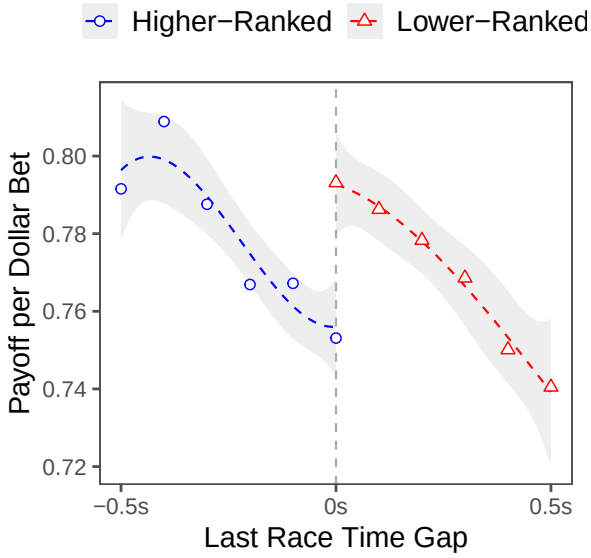
(A) Payoffs for Winning Bets by Ranking



(B) Payoffs for Placing Bets by Ranking



(C) Payoffs for Winning Bets Around Ranking Thresholds



(D) Payoffs for Placing Bets Around Ranking Thresholds

Figure 7: Ranking and Betting Payoffs

This figure examines the implications of ranking effects for bettors' payoffs using the full sample. Panels A and B plot the average gross return per dollar wagered as a function of a horse's ranking in its previous race, with the x-axis denoting ranking in the prior race. In Panels C and D, the x-axis represents the difference in race times between each horse and its adjacently ranked counterpart in its previous race. In Panels A and B, circles represent the unconditional means of payoff per dollar wagered, grouped by a horse's ranking in its previous race. In Panels C and D, circles and triangles represent the unconditional means of payoff per dollar wagered by race time gap in the previous race. The dashed lines depict fitted values from a cubic polynomial. In all panels, gray shaded areas indicate 95 percent confidence intervals. To mitigate the influence of extreme values, the analysis focuses on win and place bets placed on horses with odds below 100.

metrics—such as mutual fund star ratings, credit rating categories, and governance scores—may similarly mislead investors and distort decision-making. These results also carry policy implications regarding how information should be presented. While discretized information can reduce cognitive burden, they may also introduce perceptual biases. It is therefore essential to weigh the benefits of information discretization against its potential to distort perceptions.

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*.

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.

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

Chen, Y., Kumar, A. and Zhang, C. (2021) Searching for Gambles: Gambling Sentiment and Stock Market Outcomes, *Journal of Financial and Quantitative Analysis*, **56**, 2010–2038.

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.

Choi, J. J., Laibson, D. and Madrian, B. C. (2009) Mental Accounting in Portfolio Choice: Evidence from a Flypaper Effect, *American Economic Review*, **99**, 2085–2095.

Choi, J. J., Laibson, D. and Madrian, B. C. (2010) Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds, *The Review of Financial Studies*, **23**, 1405–1432.

Choi, J. J. and Roberson, A. Z. (2020) What Matters to Individual Investors? Evidence from the Horse's Mouth, *The Journal of Finance*, **75**, 1965–2020.

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.

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

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.

Friedman, M. and Savage, L. J. (2024) The utility analysis of choices involving risk1, in *The Foundations of Price Theory Vol 4*, Routledge, pp. 135–162.

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*.

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.

Griffith, R. M. (1949) Odds adjustments by american horse-race bettors, *The American Journal of Psychology*, **62**, 290–294.

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. (2015) The Worst, the Best, Ignoring All the Rest: The Rank Effect and Trading Behavior, *The Review of Financial Studies*, **28**, 1024–1059.

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.

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.

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*.

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

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.

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.

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	49.66	45.80	13.73	40.16	57.43
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	228,135	0.19	0.00	0.40	0.00	0.00
Male	228,135	0.29	0.00	0.45	0.00	1.00
Female	228,135	0.52	1.00	0.50	0.00	1.00
Age	228,135	3.67	3.00	1.20	3.00	4.00
Days from Last Race	228,135	31.60	28.00	20.46	21.00	35.00
Change in Tier	228,135	0.00	0.00	0.00	0.00	0.00
Higher-Ranked in Last Race	228,135	0.50	1.00	0.50	0.00	1.00
Last Race Time Gap	228,135	−0.00	0.00	0.25	−0.20	0.20
Last Race Distance Gap ≤ Nose	228,135	0.06	0.00	0.23	0.00	0.00
1 + Win Betting Odds	228,135	99.44	38.20	567.76	15.90	85.40
1 + Place Betting Odds	228,065	29.18	6.10	443.97	3.20	12.00
Win	228,135	0.04	0.00	0.20	0.00	0.00
Place	228,135	0.17	0.00	0.38	0.00	0.00
Race Characteristics						
Age-Sex Based Handicapping	228,135	0.51	1.00	0.50	0.00	1.00
Award Based Handicapping	228,135	0.49	0.00	0.50	0.00	1.00
Race Distance (meter)	228,135	1248.63	1200.00	273.01	1000.00	1400.00
N. of Horses	228,135	11.29	12.00	1.90	10.00	12.00
Seoul Stadium	228,135	0.65	1.00	0.48	0.00	1.00
Jeju Stadium	228,135	0.16	0.00	0.36	0.00	0.00
Busan Stadium	228,135	0.19	0.00	0.39	0.00	0.00
Avg. Bet Size (1K KRW)	183,751	52.64	49.18	13.30	43.30	59.80
Sales (1B KRW)	221,525	3.22	3.09	2.10	1.75	4.57

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. The exchange rate is typically 1,200 KRW per 1 USD. 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	6,527	0.19	0.39	6,537	0.19	0.39	−0.00	(−0.16)
Male	6,527	0.29	0.45	6,537	0.30	0.46	−0.01	(−1.32)
Female	6,527	0.52	0.50	6,537	0.51	0.50	0.01	(1.33)
Age	6,527	3.68	1.18	6,537	3.67	1.17	0.00	(0.12)
Horse Weight	6,527	426.25	71.88	6,537	426.79	71.59	−0.54	(−0.43)
ΔHorse Weight	6,527	−0.34	6.90	6,537	−0.20	6.71	−0.14	(−1.17)
Days from Last Race	6,527	31.23	19.80	6,537	30.96	19.53	0.27	(0.78)
Race Characteristics								
Race Distance (meter)	6,527	1253.48	270.36	6,537	1251.20	267.96	2.28	(0.48)
N. of Horses	6,527	11.29	1.92	6,537	11.30	1.92	−0.01	(−0.26)
Sunny Day	6,527	0.73	0.45	6,537	0.73	0.44	−0.01	(−0.86)
Gate Assignment								
Gate 1	6,527	0.09	0.28	6,537	0.09	0.29	−0.01	(−1.35)
Gate 2	6,527	0.08	0.26	6,537	0.07	0.26	0.00	(1.00)
Gate 3	6,527	0.05	0.21	6,537	0.05	0.22	−0.00	(−0.91)
Gate 4	6,527	0.04	0.20	6,537	0.04	0.20	−0.00	(−0.02)
Gate 5	6,527	0.02	0.13	6,537	0.02	0.14	−0.00	(−0.75)
Gate 6	6,527	0.01	0.12	6,537	0.01	0.11	0.00	(0.61)
Gate 7	6,527	0.10	0.30	6,537	0.09	0.29	0.01	(1.59)
Gate 8	6,527	0.09	0.29	6,537	0.09	0.28	0.01	(1.10)
Gate 9	6,527	0.09	0.29	6,537	0.09	0.29	0.00	(0.97)
Gate 10	6,527	0.09	0.29	6,537	0.10	0.29	−0.01	(−1.39)
Gate 11	6,527	0.10	0.29	6,537	0.10	0.30	−0.00	(−0.36)
Gate 12	6,527	0.09	0.28	6,537	0.09	0.29	−0.00	(−0.19)
Gate 13	6,527	0.08	0.28	6,537	0.08	0.28	0.00	(0.28)
Gate 14	6,527	0.07	0.26	6,537	0.08	0.27	−0.00	(−0.80)

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 (8). 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 with a time gap of 0.5 seconds or less, 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 (7)). 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 heteroscedasticity-consistent standard errors. $*$, $**$, 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	0.0161	0.0416	0.0060	0.0125
	(0.53)	(1.43)	(0.35)	(0.75)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,064	11,045	228,135	192,326
Adj. R^2	-0.00	0.20	0.00	0.21
Mean of Dep. Var.	53.94	53.95	53.92	53.91

(B) Jockey Ability

Dep. Var.:	*Jockey Ability*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.0007	0.0034	−0.0016	−0.0009
	(0.06)	(0.38)	(−0.25)	(−0.17)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,057	11,031	227,971	192,058
Adj. R^2	-0.00	0.40	0.00	0.40
Mean of Dep. Var.	6.13	6.15	6.14	6.15

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 with a time gap of 0.5 seconds or less, 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 (7)). 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 heteroscedasticity-consistent standard errors. $*$, $**$, 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	−0.3590	−0.3924	−0.1442	−0.1117
	(−0.99)	(−1.02)	(−0.70)	(−0.50)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,064	11,045	228,135	192,326
Adj. R^2	-0.00	0.03	0.00	0.01
Mean of Dep. Var.	4.49	4.40	4.25	4.23

(B) Place

Dep. Var.:	*Place (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.5226	0.4226	0.0279	0.0668
	(0.77)	(0.59)	(0.07)	(0.16)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,064	11,045	228,135	192,326
Adj. R^2	-0.00	0.06	0.00	0.06
Mean of Dep. Var.	18.63	18.90	17.25	17.39

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 with a time gap of 0.5 seconds or less, 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 (7)). 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 heteroscedasticity-consistent standard errors. $*$, $**$, 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	0.3904***	0.4514***	0.2591***	0.3149***
	(3.59)	(4.19)	(4.13)	(5.02)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,064	11,045	228,135	192,326
Adj. R^2	0.00	0.16	0.00	0.15
Mean of Dep. Var.	4.57	4.53	4.30	4.28

(B) Placing Bet

Dep. Var.:	*Bet to Place (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.3486***	0.3905***	0.2891***	0.3267***
	(3.88)	(4.55)	(5.60)	(6.58)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	0.5s	0.5s
Obs	13,058	11,035	228,065	192,230
Adj. R^2	0.00	0.23	0.01	0.22
Mean of Dep. Var.	6.39	6.39	6.10	6.10

Table 7: Are Results Driven by Next-Race Competition?

This table extends the analyses presented in Tables 5 and 6 by incorporating next-race fixed effects to account for potential differences in the competitive environments faced by higher- and lower-ranked horses, such as variation in the ability levels of their future competitors. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." All regression specifications are estimated using pairs with a time gap of 0.5 seconds or less and include pair fixed effects. In addition, they 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 $D_{r,n}$, an indicator variable for higher-ranked horses (see Equation (7)). The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using heteroscedasticity-consistent standard errors. $*$, $**$, 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.:	*Win (%)* (1)	*Place (%)* (2)	*Bet to Win (%)* (3)	*Bet to Place (%)* (4)
Higher-Ranked	−0.0097	−0.2860	0.3808***	0.4076***
	(−0.04)	(−0.56)	(5.39)	(7.28)
Pair FEs	✓	✓	✓	✓
Next Race FEs	✓	✓	✓	✓
Higher×Cubic	✓	✓	✓	✓
Sample: $\Delta \leq$	0.5s	0.5s	0.5s	0.5s
Obs	182,034	182,034	182,034	181,940
Adj. R^2	0.08	0.12	0.30	0.38
Mean of Dep. Var.	4.18	17.24	4.22	6.03

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 with a time gap of 0.5 seconds or less, 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 (7)). All regression specifications include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using heteroscedasticity-consistent standard errors. $*$, $**$, 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	0.0834	0.1133	0.0590	0.1314**
	(0.36)	(1.42)	(0.31)	(2.04)
Higher-Ranked × 2000s	0.5310***	0.3604***	0.4439***	0.3503***
	(2.92)	(5.07)	(3.14)	(6.22)
Higher-Ranked × 2010s	0.5943***	0.3635***	0.5245***	0.3999***
	(3.28)	(5.12)	(3.67)	(7.06)
Higher-Ranked × 2020s	0.3124	0.3322***	0.3414	0.3234***
	(0.86)	(2.93)	(1.11)	(3.46)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	13,064	228,135	13,058	228,065
Adj. R^2	0.16	0.15	0.23	0.22
Mean of Dep. Var.	4.53	4.28	6.39	6.10

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, across different bet size categories: Low, Middle, and High, based on the distribution of average bet size within each racecourse-year. 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 30 percent of the distribution for that racecourse in that year, Middle Bet Size equals one if it falls in the middle 40 percent, and High Bet Size equals one if it falls in the top 30 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 with a time gap of 0.5 seconds or less, 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 (7)). All regression specifications include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using heteroscedasticity-consistent standard errors. $*$, $**$, 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	0.8065***	0.5139***	0.6127***	0.4512***
	(3.52)	(3.75)	(3.24)	(4.05)
Higher-Ranked × Middle Bet Size	0.4825**	0.3796***	0.3941**	0.4264***
	(2.01)	(2.82)	(2.15)	(4.07)
Higher-Ranked × High Bet Size	0.4121	0.2840*	0.4303**	0.2602**
	(1.56)	(1.82)	(2.09)	(2.19)
Pair FEs	✓	✓	✓	✓
Bet Size FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	8,954	154,996	8,950	154,946
Adj. R^2	0.15	0.14	0.22	0.21
Mean of Dep. Var.	4.62	4.39	6.51	6.24

Table 10: Does the Magnitude of the Effects Vary Across Ranking Thresholds?

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 the horse's ranking in the previous race. The variable Higher-Ranked (ith) is an indicator equal to one if the higher-ranked horse in the pair is ranked ith in its race. 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 with a time gap of 0.5 seconds or less, 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 (7)). All regression specifications include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using heteroscedasticity-consistent standard errors. $*$, $**$, 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 (6th)	0.5329**	0.6178***	0.6845***	0.6483***
	(2.18)	(4.20)	(3.79)	(6.03)
Higher-Ranked (7th)	0.8115***	0.3814***	0.4470**	0.3183***
	(3.51)	(2.89)	(2.41)	(3.02)
Higher-Ranked (8th)	0.0126	0.1528	−0.0328	0.1944*
	(0.06)	(1.11)	(−0.17)	(1.71)
Higher-Ranked (9th)	0.4126*	0.0927	0.3522*	0.1380
	(1.66)	(0.65)	(1.70)	(1.13)
Higher-Ranked (10th)	0.2988	0.1612	0.2962	0.1257
	(0.96)	(1.03)	(1.18)	(0.93)
Higher-Ranked (11th)	0.2314	0.0136	0.1556	0.0540
	(0.68)	(0.08)	(0.50)	(0.32)
Higher-Ranked (12th)	0.6262	0.0227	0.4529	0.1205
	(1.25)	(0.08)	(1.12)	(0.46)
Higher-Ranked (13th)	0.0324	0.0193	0.9046	0.3726
	(0.04)	(0.06)	(1.28)	(0.98)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	13,064	228,135	13,058	228,065
Adj. R^2	0.16	0.15	0.23	0.22
Mean of Dep. Var.	4.53	4.28	6.39	6.10

Table 11: Implications for Betting Strategies

This table compares the realized payoff per dollar bet on higher- and lower-ranked horses using the full sample. To reduce the influence of extreme values, Panel A restricts the analysis to win and place bets on horses with odds below 50. Panel B expands the cutoff to odds below 100, while Panel C further extends it to odds below 200. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." 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 with a time gap of 0.5 seconds or less, 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 (7)). All regression specifications include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using heteroscedasticity-consistent standard errors. $*$, $**$, and $***$ indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Odds Cutoff of 50

Dep. Var.:	*Payoff per Dollar Bet to Win*		*Payoff per Dollar Bet to Place*	
	(1)	(2)	(3)	(4)
Higher-Ranked	-0.0951^{***}	-0.0521^{***}	-0.0381^{**}	-0.0236^{***}
	(-2.74)	(-5.24)	(-2.37)	(-5.04)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	36,144	754,966	51,376	1,126,276
Adj. R^2	0.00	-0.00	0.01	0.00
Mean of Dep. Var.	0.77	0.79	0.78	0.77

(B) Odds Cutoff of 100

Dep. Var.:	*Payoff per Dollar Bet to Win*		*Payoff per Dollar Bet to Place*	
	(1)	(2)	(3)	(4)
Higher-Ranked	-0.0688^{*}	-0.0367^{***}	-0.0330^{**}	-0.0240^{***}
	(-1.86)	(-3.49)	(-2.00)	(-5.02)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	45,117	966,830	51,562	1,132,613
Adj. R^2	-0.00	-0.00	0.01	0.00
Mean of Dep. Var.	0.76	0.77	0.78	0.77

Table 11: Implications for Betting Strategies (Continued)

(C) Odds Cutoff of 200

Dep. Var.:	*Payoff per Dollar Bet to Win*		*Payoff per Dollar Bet to Place*	
	(1)	(2)	(3)	(4)
Higher-Ranked	-0.0477	-0.0196^{*}	-0.0328^{**}	-0.0245^{***}
	(-1.31)	(-1.76)	(-1.99)	(-5.03)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.5s	Nose	0.5s
Obs	49,690	1,083,151	51,578	1,133,097
Adj. R^2	-0.00	0.00	0.01	0.00
Mean of Dep. Var.	0.73	0.74	0.78	0.77

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 (8).
- **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.
- **Payoff per Dollar Bet to Win**: Realized payoff per dollar wagered on a horse to win.
- **Payoff per Dollar Bet to Place**: Realized payoff per dollar wagered on a horse to place.