


## ORIGINAL ARTICLE

# Racial Discrimination in Asset Prices: Evidence From Horse Betting

Spencer Barnes<sup>1</sup> | Luke C. D. Stein<sup>2</sup> 

<sup>1</sup>University of Texas at El Paso, El Paso, Texas, USA | <sup>2</sup>Babson College, Wellesley, Massachusetts, USA

**Correspondence:** Spencer Barnes ([scbarnes@utep.edu](mailto:scbarnes@utep.edu))

**Revised:** 24 October 2025 | **Accepted:** 18 November 2025

## ABSTRACT

In the presence of behavioral biases, prices can diverge from fundamentals, and the effects of racial/ethnic bias are evident in many financial and nonfinancial markets. We investigate the determinants and consequences of discrimination in parimutuel horse betting by assessing return differences across horses whose trainers have racially/ethnically distinctive surnames, which bettors may see as a proxy for quality (accurately or inaccurately) or a source of nonpecuniary returns (due to animus). Bets on horses with non-White-named trainers earn higher risk-adjusted returns, and these differences are especially pronounced among riskier bets, which receive lower average returns under the well-known “favorite–longshot” bias. Racial/ Ethnic return differences are stronger—overall and especially among longshots—for “win” than “place” and “show” bets, among horses with poor prior performance, and in low-stakes races with “fast” conditions. These results are consistent with the effects of discrimination being strongest among less-informed and less-sophisticated bettors.

**JEL Classification:** G14, G41, J15, J71, L83, Z20, Z23

*“You wake up and you’re like, ‘I’ve got to take care of my horse.’ And the horse doesn’t get taken care of [any better] than any Caucasian trainer would take care of it, but because of their skin color, they get pushed to the top. . . . It does go back to racism.”* —Uriah St. Lewis, Jr., Horse trainer

## 1 | Introduction

Discrimination pervades many aspects of society, often leading to significant disparities in various markets, both financial and nonfinancial. Many financial assets are connected to individuals, and discrimination against these individuals can cause systematic divergence between prices and fundamental values. In this paper, we use a sports gambling market to help understand how and

why racial and ethnic characteristics affect investor behavior and, therefore, equilibrium pricing.

In particular, we investigate the risk–return relationships for bets placed on thoroughbred horses as a function of the perceived race/ethnicity of the horses’ trainers. We find systematic differences in returns consistent with discrimination: Horses with non-White-named trainers earn higher realized returns, suggesting that bettors avoid these horses. These differences are most pronounced for bets, horses, and horseraces likely to attract less-informed or less-sophisticated bettors, and robust to inclusion of horserace, jockey, and trainer fixed effects. This is consistent with limited arbitrage by sophisticated investors precluding the effects of discrimination from getting “competed away” in some markets (as in Becker, 1971).

Horse racing revolves around betting and represents a considerable portion of betting markets in the United States with around \$12 billion bet in 2022, or about 14% of the total wagered by

U.S. sports bettors (Fenly, 2023; Yakowicz, 2023). Horses race throughout the United States, and on a typical peak-season (May–June) Saturday, around 30 tracks each hold 8–12 races with an average of nine horses running in each one. This sport is deeply ingrained in American culture and presents a useful setting to examine the causes and consequences of biases due to features both of sports betting markets generally and of horse racing in particular.

As with many other betting markets, horse racing lacks exposure to systematic risk and offers known, short termination windows, allowing mispricing to surface quickly.<sup>1</sup> Unlike with most other sports, bets on horseraces are parimutuel, meaning that gamblers bet against each other (rather than a bookmaker, casino, or other “house”), with equilibrium odds determined simply by the amounts bet on each horse, and payouts determined by these odds and the horserace outcome. Legal, parimutuel betting means that unlike in many other gambling contexts, payouts for every bet are publicly posted, allowing bettors and researchers to observe the returns associated with the market’s equilibrium prices. This is a setting in which the data to identify systematic mispricing—including as a consequence of discrimination—is available for many distinct assets, but there are significant limits to arbitrage that may make it difficult for even well-informed, well-capitalized bettors to compete mispricing away.<sup>2</sup>

Bettors at the track have access to a wealth of information about horses, jockeys, and trainers, and can observe the evolution of equilibrium odds as price discovery occurs. Highly salient information relevant to assessing both “operating” and financial prospects is posted at the track (and online), and distributed through race cards and programs. Several examples are illustrated in Figures 1 and IA.1.

Among the prominent facts is the identity of trainers, who are rightly understood to be an important determinant of their horses’—and therefore bettors’—success. Trainers not only oversee horses’ day-to-day operations but are also responsible for planning workouts, entering horses in appropriate races, advising jockeys on strategy, notifying owners of the progress of their horses’ training and race entry options, supervising stable employees, and scheduling health care and maintenance appointments. Much as an investor might assess the prospects of a firm together with its CEO and board, or a fund together with its managers, bettors assess the prospects of a horse together with its trainer.

Given trainers’ importance to performance, bettors *should* pay attention to trainer quality in deciding which horses to back. However, given limited direct information, bettors may rely on trainer attributes that they believe (accurately or otherwise) are related to quality.<sup>3</sup> This could include race or ethnicity, which bettors presumably infer—consciously or not—from trainers’ surnames, which are visible and highly salient.<sup>4</sup>

Some bettors may also have nonpecuniary reasons for preferring horses with trainers from a particular demographic group. For example, a bettor might enjoy her winnings more when they represent shared success with a horse’s White trainer, or serve to support her pre-existing belief in White trainers’ superiority. Although we cannot directly observe bettors’ assessments of

quality or preferences over trainers (perhaps reflecting animus)—nor can we cleanly distinguish statistical from taste-based discrimination—we can learn something about these phenomena by examining patterns of racialized differences in returns.<sup>5</sup>

We analyze bets on 74,988 entrants in 9164 U.S. horseraces held between 2011 and 2022, focusing on the relationship between parimutuel odds, realized returns, and trainers’ surnames. We assess the extent to which a trainer might be perceived by bettors as likely to be non-White using the frequency of surnames in the 2010 U.S. Decennial Census. Horses with non-White-named trainers earn higher realized returns, consistent with bettors avoiding these horses; in the parlance of the track, these horses are “under-bet.” We also confirm that non-White-named trainers’ horses finish further ahead in the field and win more often than would be predicted by their odds, consistent with bettor discrimination against these trainers.

Of course, bets differ not only in their returns, but also in their levels of risk. Even though horseracing outcomes are entirely idiosyncratic, our analysis needs to take place in the context of the “favorite-longshot” bias, or FLB: the longstanding empirical regularity that betting odds give inaccurate estimates of the likelihood that a horse will win (see, e.g., Thaler and Ziemba, 1988). In particular, favorites are consistently under-bet (generating higher expected returns) while longshots are over-bet (generating lower expected returns). We find that non-White-named trainers generate *higher* returns despite the fact that they are somewhat more likely to train (*low-expected-return*) longshots. That is, risk-adjustment not only fails to eliminate the racial/ethnic return gap, but, if anything, makes it slightly larger.

A variety of neoclassical and behavioral models can help explain the existence of the FLB. Key ingredients in many explanations include bettors with risk-seeking preferences (or perhaps a “recreational interest” in gambling, as in Ottaviani and Sørensen, 2010) or heterogeneous beliefs about horses’ prospects (as in Gandhi and Serrano-Padial, 2015). Lower average returns on longshots disproportionately attract less-informed bettors and those with nonpecuniary, recreational tolerance for risk-taking (Feess et al., 2014; Geertsema and Schumacher, 2016). These gamblers should therefore play a particularly important role in pricing longshot assets. If racial/ethnic return gaps are larger for longshots, the effects of discrimination are likely strongest among less-informed and less-sophisticated bettors.

Our key finding that horses with non-White-named trainers earn higher realized returns is especially pronounced among longshots. That is, the FLB is more severe for horses with White-named trainers. This holds even in estimates with trainer fixed effects, which implicitly compare returns on the same trainers’ horses when they race as favorites versus longshots. Our results are consistent with the idea that avoiding non-White-named trainers due to discrimination is most pronounced among less-informed and less-sophisticated bettors.

We also assess whether these return differences—both overall and heterogeneously across risk levels—vary across several bet, horse, and horserace characteristics. The effects are larger among “win” bets (where a bettor has to pick the winner) than “place” and “show” bets (which succeed if a horse finishes in the top

**RACE 7 - POST TIME - 3:54 PM CT**

ENTRIES PLUS (FREE)

SMART PICK (FREE)

PP (RACE 7)

Daily Double / Exacta / 50 Cent Trifecta / 10 Cent Superfecta 50 Cent Pick 3 (Races 7-8-9)

Oaklawn Park CLAIMING \$80,000 – \$50,000

Purse \$60,000. For Three Year Olds Which Have Never Won Two Races. Weight, 125 Lbs. (Horses Which Have Never Won A Race At A Mile Or Over Allowed 3 Lbs.) Claiming Price \$80,000, For Each \$10,000 To \$50,000 3 Lbs. One Mile.

| P# | PP | Horse ?                               | Virtual Stable           | A/S | Med | Claim \$ | Jockey                         | Wgt | Trainer                      | M/L  |
|----|----|---------------------------------------|--------------------------|-----|-----|----------|--------------------------------|-----|------------------------------|------|
| 1  | 1  | <a href="#">Accidental Hero (KY)</a>  | <input type="checkbox"/> | 3/C | L   | \$70,000 | <a href="#">F Arrieta</a>      | 122 | <a href="#">S M Asmussen</a> | 12/1 |
| 2  | 2  | <a href="#">Judo (KY)</a>             | <input type="checkbox"/> | 3/C | L   | \$80,000 | <a href="#">R Santana, Jr.</a> | 125 | <a href="#">B H Cox</a>      | 3/1  |
| 3  | 3  | <a href="#">Giroovin (FL)</a>         | <input type="checkbox"/> | 3/C | L   | \$80,000 | <a href="#">I Castillo</a>     | 122 | <a href="#">T W Fincher</a>  | 7/2  |
| 4  | 4  | <a href="#">Sir (KY)</a>              | <input type="checkbox"/> | 3/C | L   | \$50,000 | <a href="#">W De La Cruz</a>   | 116 | <a href="#">C Contreras</a>  | 6/1  |
| 5  | 5  | <a href="#">Tres Soles (KY)</a>       | <input type="checkbox"/> | 3/C | L   | \$70,000 | <a href="#">C A Torres</a>     | 119 | <a href="#">S M Asmussen</a> | 8/1  |
| 6  | 6  | <a href="#">Recker Point (KY)</a>     | <input type="checkbox"/> | 3/C | L   | \$80,000 | <a href="#">M Murrill</a>      | 125 | <a href="#">C A Hartman</a>  | 8/5  |
| 7  | 7  | <a href="#">G T Five Hundred (KY)</a> | <input type="checkbox"/> | 3/G | L   | \$80,000 | <a href="#">T Wales</a>        | 122 | <a href="#">I Mason</a>      | 12/1 |

**Owners:** 1 - [West Point Thoroughbreds \(Terrance Finley, Edwin S. Baker and William Sandbrook\)](#) ; 2 - [Gary & Mary West Stables \(Gary & Mary West\)](#) ; 3 - [Triple V Racing LLC \(Brandon & Sarah Valentini\)](#) ; 4 - [Flurry Racing Stables LLC \(Staton Flurry\)](#) ; 5 - [Winchell Thoroughbreds LLC \(Ron & Joan Winchell\)](#) ; 6 - [JD Thoroughbreds LLC \(Jackie Slawson, Joe K. Davis & Larry Romero\)](#) ; 7 - [Muddy Waters Stables LLC \(Mike Waters\)](#)

**Breeders:** 1 - Michael T. Barnett; 2 - Gary & Mary West Stables Inc.; 3 - Bella Inizio Farm, LLC; 4 - Brad Anderson; 5 - Rock Brothers Breeding LLC; 6 - Jumping Jack Racing LLC & Seclusive Farm; 7 - Joe B. Mulholland Jr., John P.Mulholland & Karen Mulholland

**Pedigrees (Sire - Dam, by Dam Sire):** 1 - [Mo Town - Pardonmecomingthru](#), by [Chatain](#) ; 2 - [Street Sense - Our Love Tap](#), by [Tapit](#) ; 3 - [Girvin - Purecraziluck](#), by [Simon Pure](#) ; 4 - [Munnings - Taylor A. Little J](#), by [Medaglia d'Oro](#) ; 5 - [Justify - Chocolate Souffle](#), by [Tapit](#) ; 6 - [Kantharos - Uknowwhatimean](#), by [Indian Charlie](#) ; 7 - [Astern \(AUS\) - Cocoanut Row](#), by [Mineshaft](#) ;

**FIGURE 1** | Racing card. The figure shows an example of a horserace card entry from Equibase prominently displaying trainer surnames. Several other examples of information available to bettors are included in Figure IA.1.

two or three) that are particularly popular among sophisticated gamblers. Effects are also stronger among horses with poor prior performance, consistent with discrimination being more prevalent among less-informed gamblers who fail to correctly handicap persistence. We also show that racial/ethnic return differences are stronger in low-stakes races with “fast” conditions, precisely where the returns to information acquisition may be low, so betting is more dominated by less-sophisticated gamblers.

The paper contributes to several strands of literature. Our investigation of trainer race/ethnicity complements earlier work on discrimination in horseracing based on jockey demographics. <sup>6</sup> Citetbinder2021money and Binder and Grimes (2022) find that female jockeys in the U.S. underperform relative to bettor expectations, but evidence from the United Kingdom suggests female jockeys exhibit relative overperformance there (Brown

and Yang, 2015; Cashmore et al., 2022). Leeds and Rockoff (2019, 2020) show that between 1875 and 1915, African-American jockeys, who were initially numerous, were forced out of the three prestigious U.S. “triple crown” horseraces, and that bettors discriminated against these jockeys.<sup>7</sup>

It is important to understand how trainer attributes affect pricing given that trainers are highly diverse, salient to bettors, and critical in driving both operating and financial success. We also connect the discrimination and FLB literatures, explicitly considering asset returns and introducing a set of tests that assess bias across the odds distribution to better understand mechanisms underlying discrimination’s effects. Together with our other new tests exploiting variation across bet, horse, and horserace characteristics, we are able to identify where discrimination against non-White-named trainers has the largest effects.

Our approaches should also be useful to researchers investigating other forms of deviation from fair pricing in horse racing.

More broadly, we show how the financial effects of discrimination can vary even within a market.<sup>8</sup> Identifying and mitigating discrimination is easier if we understand which market features are correlated with larger adverse effects. We consider various sources of cross-sectional and time-series variation to show that the pricing effects of bias are driven by the least sophisticated or informed bettors, and market segments where limits to arbitrage prevent these effects from getting competed away. Although these specific sources of variation are of course unique to horse racing, our results suggest that discrimination may have the largest effects where limits to arbitrage are strong and investors are unsophisticated or poorly informed, hypotheses worth investigating in other markets.

## 2 | Horse Racing Primer

Among betting markets, horse races provide one of the closest analogs to more traditional financial asset markets. In the United States, the specific structure is parimutuel betting. A bettor puts their money on a horse to win. All these win bets are placed in a pool, with the amount of money bet on each horse used to calculate its odds. Bettors who chose the winning horse share the pool pro rata, minus a “takeout” retained by the track.<sup>9</sup>

Payoffs do not depend on the odds posted *when* a bet is placed, but rather only on the total amount bet *as of the time the race starts*. These “starting odds” are determined by the demand for bets on each horse and, therefore, influenced by any factors that affect this demand. These include horses’ past performance, names, or numbers; jockeys’ identity or the color of their racing silks; track conditions; and other considerations.

However, one of the most critical factors that bettors may use to assess a horse is its trainer. Analogous to the role a CEO plays at a firm, a trainer oversees not only the day-to-day operation of the horse but also makes executive decisions on how to make the horse perform at its peak. Specifically, trainers are responsible for planning workouts, entering horses in appropriate races, advising the jockey on race strategy, notifying owners of the progress of their horses’ training and race entry options, supervising stable employees, and scheduling health care and maintenance appointments. Almost everything is in the hands of the trainer.<sup>10</sup>

Figure 1 shows an example of a race card entry, containing standard information readily available to all bettors. Race seven at Oaklawn Park in Hot Springs, Arkansas, was set to go off at 3:54 p.m., offering win, place, and show bets, along with a variety of “exotic” bets. The card shows that the purse—prize money received by the horses’ owners, trainers, and jockeys—is \$60,000, and provides information on the horses, jockeys, owners, breeders, and horses’ pedigrees. Among this information is the trainer’s surname, prominently displayed along with the horse names.

We use these surnames to assess what bettors might infer about trainers’ race/ethnicity. While a small number of the most famous trainers may be well known among bettors, there are thousands of

active trainers, and demographic signals other than their names are not typically highly salient. In addition to trainer surnames being prominently shown on online racing cards, we show in Figure 1A.1 several other places where they are visible to bettors when placing bets at an electronic terminal, on their phones, or while consulting a paper racing card. Whether or not bettors are intentionally assessing trainers’ races or ethnicities, surnames are salient and can trigger stereotypes (explicit or implicit) or other forms of bias.

## 3 | Data

### 3.1 | Data Collection

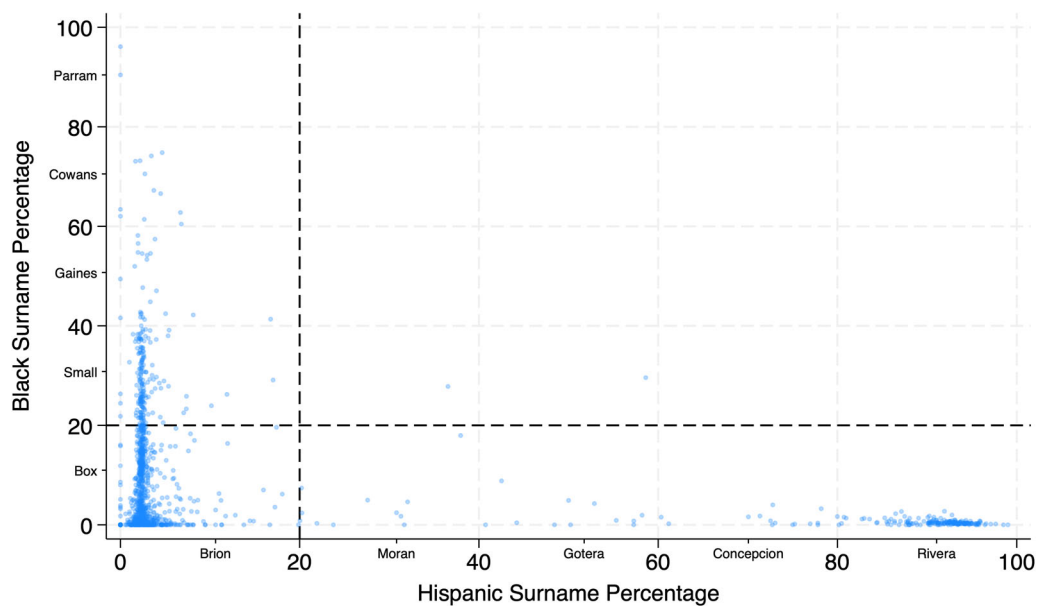
Our horse racing data comes from web-scraping publicly available U.S. horse racing log websites without sign-ups, logins, or other restrictions. Our main analysis sample is built using information from [racingpost.com](https://www.racingpost.com), which overwhelmingly covers U.S. stakes races between 2011 and 2022, comprising 9164 horseraces with 74,988 total entrants. In some analysis, we supplement Racing Post data with morning line odds and place/show pools from [horseracingnation.com](https://www.horseracingnation.com). These data are user-generated and only go back to 2018; we match horses and horseraces across these two sources for the years where they overlap.

These sources provide information about each horserace and, for each entrant, details on the horse and trainer. The logs also provide information about race results and the starting odds for each entrant. Starting odds represent the final odds posted as soon as the race starts and determine bettors’ payouts if their chosen horse wins.

Trainers’ surnames serve as an observable characteristic that can proxy for or activate stereotypes about ethnic/racial heritage, helping to identify how bias affects pricing outcomes. Surnames are often associated with a historical location around the world, and they do not change frequently (Eggers et al., 2018). People naturally associate names with these histories, influencing their economic decision-making (Bertrand and Mullainathan, 2004).

The U.S. Census Bureau’s “Frequently Occurring Surnames from the 2010 Census” data contains a demographic breakdown of the full-count population of individuals with each of 162,253 surnames in the 2010 Decennial Census. Our Racing Post sample contain 2915 unique trainers, of whom 2634 have names in the Census file. Figure 2 plots these trainers by the fraction of the population having their surname classified as Hispanic or Black.<sup>11</sup>

We consider robustness across several different ways of assessing trainers’ perceived race/ethnicity as a discrete or continuous function of surnames population demographics. In our baseline, we focus on identifying the trainers whose names most strongly identify them as White because these names are relatively common (enough to appear in the Census file) and because a relatively small fraction of the population with the name are Hispanic or Black. Names that are less distinctively white we label as “non-White,” with separate categories for more Hispanic, Black, and ambiguous names. Our baseline categorization is based on 20% Census surname percentages, indicated by horizontal and vertical



**FIGURE 2** | Trainer surname categorization. The figure graphs 2634 trainer surnames that appear in the 2011–2022 U.S. horse racing logs and the Census Frequently Occurring Surnames file as described in Section 3.1. The horizontal and vertical axes plot Hispanic and Black surname percentages in the population, respectively, and show example trainer surnames. Horizontal and vertical lines correspond to the 20% thresholds used for our baseline surname categorization. Of 2915 total trainers, 1608 are categorized as White, 465 as Hispanic, 557 as Black, and 285 as ambiguous (including 281 trainers who appear in the horse racing logs but not the Census file).

lines in Figure 2 that divide trainers into these four categories. In addition, the 281 unplotted trainers whose uncommon surnames mean they do not appear in the Census file are also categorized as ambiguous since bettors may be less likely to confidently infer whiteness from an unfamiliar name.

### 3.2 | Summary Statistics

Table 1 shows summary statistics for several key horserace and bet attributes and outcomes, as well as our baseline categorization of trainers’ surnames. Non-White-named trainers account for 35% of horse-race level observations (8% Hispanic, 15% Black, 12% ambiguous). The median finishing position is fifth place (mean 4.9), necessarily in the middle of the pack among the median horserace’s nine runners (mean 8.9). About 12% of horses win a horserace, so the median win bet does not pay out and earns a realized return of  $-100\%$ .<sup>12</sup>

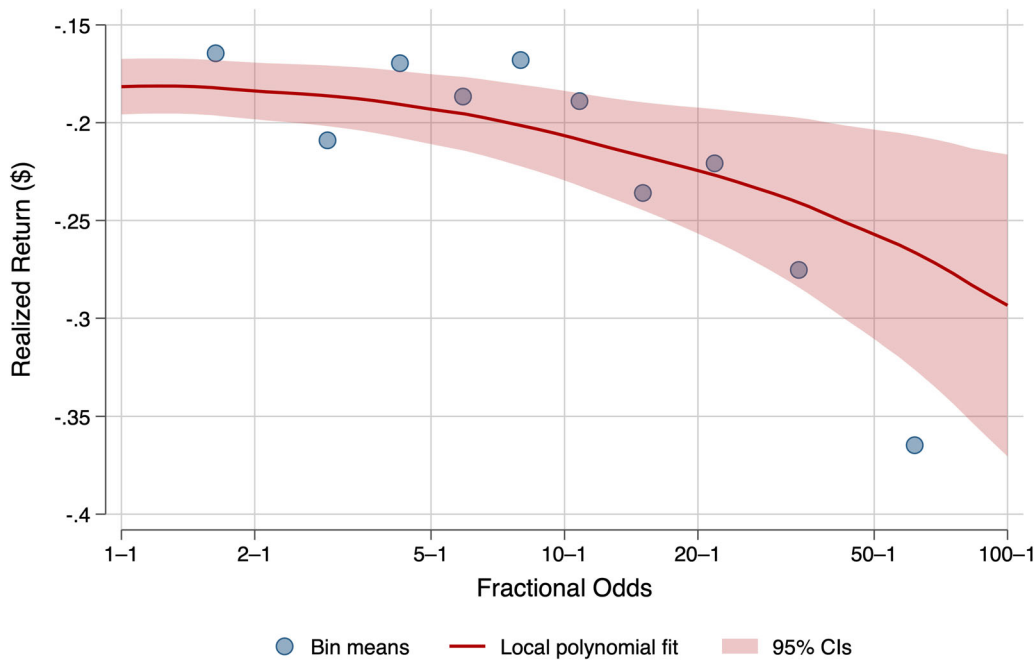
Decimal odds are the fraction (net of the track’s take, and in some cases rounded) of the parimutuel pool bet on a given horse, which we can think of as representing a market-implied risk-neutral win probability if markets were efficient. Table 1 shows that average decimal odds in our sample are 0.15; betting is on average a losing proposition, given the mean (actuarially fair) win probability of approximately 12%. We also consider an “odds-predicted position,” ranking horses within each race by their odds. Since there are the same number of horses running a horserace as the number of horses to place bets on, this measure has the same mean (and nearly equal other summary statistics) as the actual finishing position.<sup>13</sup>

The maximum potential (gross) payout for a \$1 win bet on a given horse can be calculated as  $WinAmount = 1/DecimalOdds$ . Odds

**TABLE 1** | Summary statistics.

|                                 | (1)<br>Mean | (2)<br>SD | (3)<br>Median |
|---------------------------------|-------------|-----------|---------------|
| <b>Horserace/Bet outcomes</b>   |             |           |               |
| Realized return (\$)            | -0.22       | 3.50      | -1.00         |
| Position                        | 4.90        | 2.88      | 5.00          |
| Won                             | 0.12        |           |               |
| <b>Odds</b>                     |             |           |               |
| Decimal odds                    | 0.15        | 0.15      | 0.10          |
| Fractional odds                 | 17.60       | 22.88     | 8.90          |
| Fractional odds (log)           | 2.19        | 1.22      | 2.19          |
| Odds-predicted position         | 4.93        | 2.89      | 5.00          |
| <b>Trainer surname category</b> |             |           |               |
| White                           | 0.65        |           |               |
| Non-White                       | 0.35        |           |               |
| Hispanic                        | 0.08        |           |               |
| Black                           | 0.15        |           |               |
| Ambiguous                       | 0.12        |           |               |
| <b>Horserace attributes</b>     |             |           |               |
| Runners                         | 8.85        | 2.40      | 9.00          |
| Purse (\$)                      | 181,894     | 359,531   | 96,153        |

Note: The table reports sample means, standard deviations, and medians. (For binary indicator variables, only means are reported.) Throughout the table, an observation is a horse that started in a horserace ( $N = 74,988$ ). The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.



**FIGURE 3** | Realized returns by odds. The figure graphs a local polynomial fit (and its 95% confidence interval) for realized returns from \$1 win bets and fractional odds on a log odds scale with a bandwidth of one log odds. The data are from 2011 to 2022 U.S. horse racing logs as described in Section 3.1. Variable definitions are provided in Table A.1.

are typically posted at U.S. tracks in terms of maximum potential net returns,  $FractionalOdds = WinAmount - 1$ , and generally expressed as a ratio. For example, a horse posted as “3–1” offers a maximum potential net return of 300%, a maximum potential gross payout of \$4 on a \$1 bet, and decimal odds of 0.25.

We follow Green et al. (2020) to measure bet performance, assessing realized (net) return as a function of  $WinAmount$  and race results, using an indicator for whether the horse won the race ( $Won$ ), accounting for the rare occurrence of “dead heats” (effectively ties, characterized by  $Winners > 1$ ) where the parimutuel pool is shared across bets on multiple horses:  $RealizedReturn = WinAmount \cdot (Won/Winners) - 1$ .

Table 1 shows that the average realized return in the sample is –\$0.22.<sup>14</sup> Realized returns will serve as the key dependent variable in our empirical analyses.

## 4 | Empirical Approach and Results

### 4.1 | Realized Returns by Odds (Favorite–Longshot Bias)

The favorite–longshot bias (FLB) refers to the well-known empirical pattern that bets with lower fractional odds (favorites) generate higher average realized returns (Thaler and Ziemba, 1988).<sup>15</sup> We document that the empirical regularity holds in our sample.

Specifically, Figure 3 shows average realized returns on equal-weighted portfolios of win bets at each decile of the odds distribution. In addition to these means, we plot a local polynomial fit across the logarithm of fractional odds and its 95%

confidence interval. Win bets at all odds levels lose money on average due to the track takeout, the proportion of each wager kept by the racecourse as a commission. However, wagers on horses with longer odds lose more money. For example, a \$1 win bet on a 1–1 favorite loses on average about –\$0.17, a smaller loss than the equal-weighted average realized return of –\$0.22. By contrast, a 30–1 longshot on average loses –\$0.27, and a 50–1 longshot loses –\$0.36. A simple implication of the FLB is that bettors could reduce their losses by wagering on favorites instead of longshots.

We model the FLB in regression form as

$$RR_{hr} = \alpha + \gamma \log(FO_{hr}) + \xi NF_{hr} + \delta(\log(FO_{hr}) \times NF_{hr}) + \omega_r + \epsilon_{hr}. \quad (1)$$

The dependent variable,  $RR_{hr}$ , is the realized return from the \$1 win bet strategy for horse  $h$  in horserace  $r$ . The independent variables include  $\log(FO_{hr})$  for log fractional starting odds;  $NF_{hr}$ , an indicator variable capturing whether a horse is a nonfavorite based on offering above-median odds; the interaction ( $\log(FO_{hr}) \times NF_{hr}$ ), which allows slopes to vary across the favorite/nonfavorite part of the distribution; and  $\omega_r$ , horserace fixed effects.  $\epsilon_{hr}$  is the econometric error term, and we report robust standard errors.

Results from OLS estimates of Equation (1) are reported in Table 2, where columns relax successive constraints that regression coefficients be zero. Column (1) reports the –0.22 average realized return of the sample (statistically significant at the 1% level). In Column (2), we control for log fractional odds, effectively estimating a linear fit of the FLB as illustrated in Figure 3. The estimated constant of –0.12 implies that a \$1 win bet on a 1–1 favorite horse (log fractional odds of 0, decimal odds of 0.5) loses 12 cents. For a one-unit increase in log fractional odds,

TABLE 2 | Realized returns by odds.

|                    | Dependent variable: Realized return (\$) |                        |                        |                        |                        |
|--------------------|--|------------------------|------------------------|------------------------|------------------------|
|                    | (1)                                      | (2)                    | (3)                    | (4)                    | (5)                    |
| Constant           | -0.2191***<br>(0.0128)                   | -0.1166***<br>(0.0200) | -0.1066***<br>(0.0238) | -0.1649***<br>(0.0111) | -0.1342***<br>(0.0198) |
| Odds (log)         |  | -0.0468***<br>(0.0126) | -0.0613***<br>(0.0194) | -0.0131<br>(0.0111)    | -0.0369**<br>(0.0162)  |
| Nonfavorite        |  |                        | 0.0437<br>(0.0376)     | 0.2887**<br>(0.1208)   | 0.2353*<br>(0.1253)    |
| Nonfavorite × Odds |  |                        |                        | -0.1068**<br>(0.0423)  | -0.0765*<br>(0.0454)   |
| Horsrace FE        |  |                        |                        |                        | ✓                      |
| N                  | 74,988                                   | 74,988                 | 74,988                 | 74,988                 | 74,988                 |

Notes: The table presents results from estimation of Equation (1). The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

there is a corresponding 4.7 percentage point decrease in realized returns.

However, the relationship in Figure 3 exhibits nonlinearity, with binned means suggesting a kink near the sample median. A similar, stark break in the slope of the odds–return relationship is illustrated by Snowberg and Wolfers (2010, Fig. 1). We therefore allow in Column (3) a break in the level of the realized returns-to-odds relationship by including an indicator variable *Nonfavorite* for horses with above-median odds. We interact this indicator with log fractional odds in Columns (4) and (5), allowing for breaks in both level and slope.<sup>16</sup> These latter two columns—without and with horserace fixed effects—suggest that for favored horses (i.e., below-median odds), the FLB is moderate. However, in the nonfavorite part of the distribution, the slope steepens by a marginally statistically significant -0.11 to -0.08.

The existence of the FLB illustrated in Table 2, and the fact that this “bias” is stronger among nonfavorites, will inform our approach for understanding the relationship between race/ethnicity and returns across the risk distribution. Lower average returns on longshots should disproportionately attract less-informed bettors and those with nonpecuniary, recreational tolerance for risk-taking. Citetfeess2014favorite and Geertsema and Schumacher (2016) show that inexperienced and unsophisticated indeed bet more on longshots. These gamblers should therefore play a particularly important role in pricing longshot assets.<sup>17</sup> If racial/ethnic return gaps are larger for longshots, the effects of discrimination are likely strongest among less-informed and less-sophisticated bettors.

#### 4.2 | Horserace and Bet Outcomes by Trainer Race/Ethnicity

Name heterogeneity and its influence on decision-making from economic agents is a common empirical strategy to measure

racial biases (Bertrand and Mullainathan, 2004; Arai and Thour-sie, 2009; Guell et al., 2015; Grumbach and Sahn, 2020). Our methodology employs the frequency of a surname occurring in a population to proxy for what a bettor might infer about the race or ethnicity of a horse trainer, given that names appear in race programs and are highly salient, while most bettors are unlikely to have access to other signals about trainers’ demographics. Trainer surnames are not systematically related to horses’ operating performance: We find no significant difference between the distribution of White and non-White-named trainers horses’ finishing positions.<sup>18</sup> We therefore focus on assessing bettor-driven differences in financial outcomes and start to consider here how surnames might impact horse pricing outcomes (Barghouthy et al., 2020; Feigenberg and Miller, 2022).

Considerable heterogeneity exists among realized returns by trainers’ perceived race/ethnicity, suggesting that it influences pricing decisions. The average realized return for \$1 win bets on White-named trainers is -\$0.26, while non-White-named trainers have better realized returns at -\$0.15. Hispanic-named trainers generate the highest average realized returns of -\$0.13, less negative (though not statistically significantly) than Black- and ambiguously named trainers’ averages of approximately -\$0.14 and -\$0.16, respectively.<sup>19</sup>

Of course, we showed in Section 4.1 that bettors earn different average returns for favorites and longshots. In particular, the higher returns earned by horses with non-White-named trainers could be generated mechanically by the FLB if these trainers disproportionately trained favorites. However, this is not the case: Non-White-named trainers are more likely to train longshot horses,<sup>20</sup> which receive lower average returns under the FLB (as shown in Figure 3 and Table 2). The unconditional racial/ethnic return differences therefore go in the opposite direction of what the FLB alone would generate.

In Columns (1) and (2) of Table 3, we estimate returns by race/ethnicity with a linear control for log fractional odds to

**TABLE 3** | Horserace/Bet outcomes by odds and trainer surname category.

|                    | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    |
|--------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                    | Realized Return        |                        | Position               |                        | Won                    |                        |
| Constant           | -0.1542***<br>(0.0235) | -0.1522***<br>(0.0235) | 2.2569***<br>(0.0186)  | 2.2554***<br>(0.0186)  | -0.0074***<br>(0.0016) | -0.0075***<br>(0.0016) |
| Non-White          | 0.1190***<br>(0.0285)  |                        | -0.0622***<br>(0.0185) |                        | 0.0081***<br>(0.0023)  |                        |
| Hispanic           |                        | 0.1534***<br>(0.0591)  |                        | -0.1011***<br>(0.0324) |                        | 0.0088**<br>(0.0038)   |
| Black              |                        | 0.1186***<br>(0.0399)  |                        | -0.0466*<br>(0.0256)   |                        | 0.0105***<br>(0.0033)  |
| Ambiguous          |                        | 0.0962**<br>(0.0421)   |                        | -0.0552**<br>(0.0276)  |                        | 0.0046<br>(0.0036)     |
| Odds (log)         | -0.0482***<br>(0.0125) | -0.0492***<br>(0.0126) |                        |                        |                        |                        |
| Predicted position |                        |                        | 0.5414***<br>(0.0034)  | 0.5417***<br>(0.0034)  |                        |                        |
| Odds (decimal)     |                        |                        |                        |                        | 0.8513***<br>(0.0104)  | 0.8517***<br>(0.0104)  |
| <i>N</i>           | 74,988                 | 74,988                 | 74,493                 | 74,493                 | 74,988                 | 74,988                 |

Notes: The table presents results from estimation of Equation (2). The dependent variable in Columns (1) and (2) is the realized return from a \$1 win bet. The dependent variable in Columns (3) and (4) is the finishing position of a horse (where 1 is first place and did-not-finish entries are dropped from the sample). The dependent variable in Columns (5) and (6) is an indicator variable that takes the value of 1 if a horse won the horserace and 0 otherwise; the resulting estimates therefore reflect a linear probability model. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

account for the existence of the FLB and the different odds distributions across race/ethnicity. This gives regression models that are variations of

$$RR_{hr} = \alpha + \beta RE_{hr} + \gamma \log(FO_{hr}) + \epsilon_{hr}, \quad (2)$$

where  $RE_{hr}$  are indicators for the racial/ethnic name categorization. Column (1) suggests that even when risk-adjusting using log fractional odds, horses with non-White-named trainers earn 11.9 percentage point higher realized returns (significant at the 1% level). Column (2) shows that these gaps are particularly large among Hispanic-named trainers, who generate returns 15 percentage points higher than White-named trainers (again, significant at the 1% level), while Black- and ambiguously named trainers also deliver higher returns than White-named trainers.

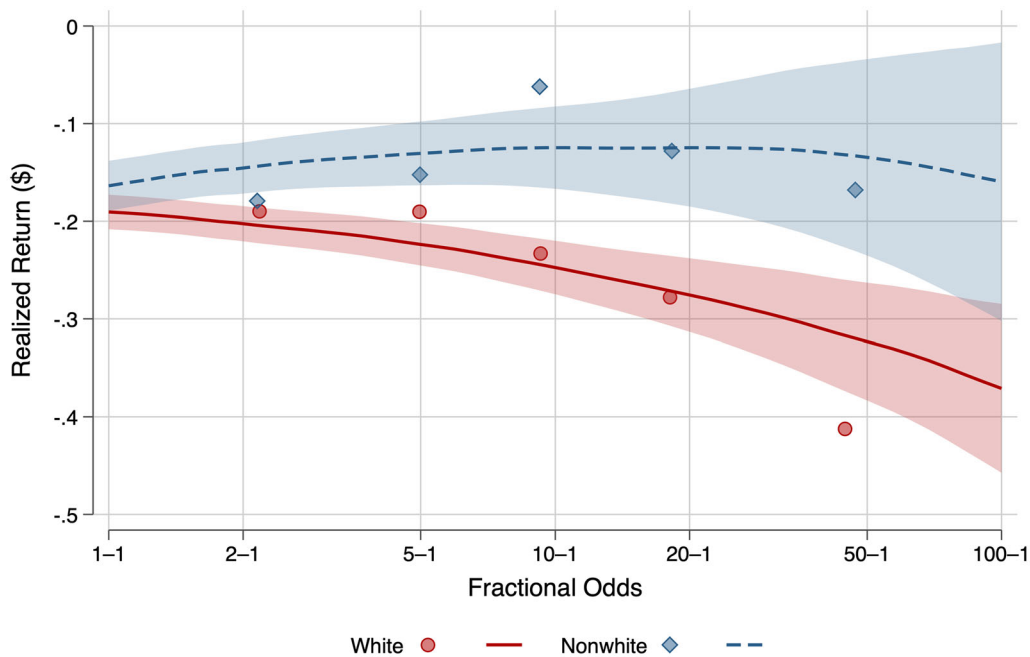
We also assess racial/ethnic gaps across two other horserace outcome measures. In Columns (3) and (4), we consider horses' actual finishing position, replacing the log fractional odds control with our measure of odds-predicted finishing position. The horse offering the lowest fractional odds is predicted to win the race, and the horse with the highest odds is predicted to come in last.

As expected, the estimated regression coefficients on predicted position are positive and significant—horses with lower odds

tend to do better—but less than one, since results are uncertain and odds are an imperfect predictor. Column (3) suggests that non-White-named trainers' horses finish on average about 0.06 spots better than a White-named-trained horse with the same odds (significant at the 1% level); Column (4) shows that this effect exists across non-White subcategories but is particularly strong for Hispanic-named trainers. These results are consistent with the effects of bettor bias appearing even among the vast majority of horserace entrants that do not wind up winning.

Finally, in Columns (5) and (6), we consider linear probability models with a binary indicator for winning as the dependent variable. We replace log fractional odds with decimal odds as a control for win likelihood (with estimated coefficients, as expected, positive but less than one). Non-White-named trainers win more horseraces than their decimal odds suggest, and Black-named trainers have the largest likelihood gaps from White-named trainers.

Taken together, the evidence presented in Table 3 shows that non-White-named trainers outperform, despite the fact that they are disproportionately likely to train longshots, who receive lower returns under the FLB. These results are consistent with underbetting of these horses, either because bettors simply prefer not to bet on non-White-named trainers or because they (inaccurately) underestimate these trainers' quality.<sup>21</sup>



**FIGURE 4** | Realized returns by odds and White/ non-White trainer surname. The figure graphs local polynomial fits (and 95% confidence intervals) for realized returns from \$1 win bets and fractional odds on a log odds scale with a bandwidth of one log odds, separately by White and non-White trainer surnames. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

### 4.3 | Realized Returns by Odds and Race/Ethnicity

Not only do *average* realized returns vary by trainer race/ethnicity, but this variation is not constant between favorites and longshots. Figure 4 plots realized returns at quintiles of the odds distribution (with a local polynomial fit and its 95% confidence interval) separately for White- and non-White-named trainers, replicating Figure 3.

White-named trainers, who comprise approximately two-thirds of the sample, exhibit a pattern consistent with traditional estimates of the FLB as in Section 4.1: Average realized returns for favorites have a nearly flat slope with respect to odds, while longer longshots offer increasingly negative returns starting near the median of the odds distribution. For example, 1-1, 5-1, and 10-1 bets with White-named trainers all average roughly -20% returns, falling to near -30% at 20-1 and -35% to -40% (local polynomial fit or binned mean) at 50-1.

The risk–return relationship looks very different for non-White-named trainers. Consistent with the mean differences discussed in Section 4.2, horses with non-White-named trainers have higher average returns, and the sign of this difference is the same at every quintile of the odds distribution. (Figure 4 also shows that confidence intervals of separate local polynomial fits are also nonoverlapping among all but the most extreme favorites and longshots.)

However, the White–non-White differences appear to be much larger among longshots than favorites. The pattern is consistent with the existence of bettor bias against non-White-named trainers’ longshots large enough to drive these horses’ returns up by

roughly 15 percentage points. The magnitude of this gap relative to the slope of the FLB suggests that bettors could reduce their losses by wagering on non-White-named trainers’ horses that are neither extreme favorites nor extreme longshots. This avoids the large negative returns on extreme longshots, while earning a premium by betting on horses underbet by biased gamblers.

We model this using ordinary least squares regressions that allow average returns and the differences in average returns across race/ethnicity to vary between favorites and longshots:

$$RR_{hr} = \alpha + \beta RE_{hr} + \gamma \log(FO_{hr}) + \xi NF_{hr} + \delta(RE_{hr} \times NF_{hr}) + \omega_r + \phi_j + \eta_t + \epsilon_{hr}. \quad (3)$$

The coefficients  $\gamma$ , on log fractional starting odds, and  $\xi$ , on an indicator for nonfavorites (i.e., horses with above median odds), capture the FLB for White-named trainers. As in Equation (2), the coefficient  $\beta$  on indicator variables for trainer name category captures realized return differences from White-named trainers, now specifically among favorites. The coefficient  $\delta$  captures *additional* racial/ethnic return differences among nonfavorites; these latter two are our key coefficients of interest. In some estimates, we include horserace fixed effects ( $\omega_r$ ), jockey fixed effects ( $\phi_j$ ), and/or trainer fixed effects ( $\eta_t$ ) to control for other time-invariant characteristics. Robust standard errors are reported.<sup>22</sup>

Results from estimating variants of Equation (3) are shown in Table 4. Column (1) serves as a baseline, including only the binary indicator for non-White-named trainer surnames; the constant (-26%) represents average realized returns for White-named trainers across the entire odds distribution. Non-White-named trainers offer an average return 12 percentage points higher.

**TABLE 4** | Realized returns by odds and White/non-White trainer surname.

|                         | Dependent variable: Realized return (\$) |                        |                        |                        |                        |                        |                        |                        |
|-------------------------|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                         | (1)                                      | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Constant                | -0.2586***<br>(0.0145)                   | -0.1542***<br>(0.0235) | -0.1438***<br>(0.0269) | -0.1146***<br>(0.0248) | -0.0983***<br>(0.0274) | -0.1532***<br>(0.0274) | -0.1556***<br>(0.0259) | -0.2200***<br>(0.0323) |
| Non-White               | 0.1151***<br>(0.0287)                    | 0.1190***<br>(0.0285)  | 0.1193***<br>(0.0285)  | 0.0423**<br>(0.0208)   | 0.0335<br>(0.0296)     | 0.0492**<br>(0.0228)   | —                      | —                      |
| Odds (log)              |  | -0.0482***<br>(0.0125) | -0.0634***<br>(0.0193) | -0.0663***<br>(0.0192) | -0.0761***<br>(0.0196) | -0.0476**<br>(0.0199)  | -0.0330*<br>(0.0200)   | -0.0059<br>(0.0220)    |
| Nonfavorite             |  |                        | 0.0459<br>(0.0376)     | -0.0010<br>(0.0420)    | 0.0097<br>(0.0463)     | -0.0096<br>(0.0429)    | -0.0271<br>(0.0426)    | -0.0225<br>(0.0481)    |
| Non-White × Nonfavorite |  |                        |                        | 0.1532***<br>(0.0565)  | 0.1699***<br>(0.0623)  | 0.1575***<br>(0.0577)  | 0.1534**<br>(0.0609)   | 0.1837***<br>(0.0687)  |
| Horseshoe FE            |  |                        |                        |                        | ✓                      |                        |                        | ✓                      |
| Jockey FE               |  |                        |                        |                        |                        | ✓                      |                        | ✓                      |
| Trainer FE              |  |                        |                        |                        |                        |                        | ✓                      | ✓                      |
| N                       | 74,988                                   | 74,988                 | 74,988                 | 74,988                 | 74,988                 | 74,689                 | 74,156                 | 73,856                 |

Notes: The table presents results from estimation of variations of Equation (3). The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

Column (2) adds the log fractional odds control (repeating Table 3, Column 1). In Column (3) we allow a break in returns between favorites and nonfavorites.

If markets were efficient and if—counterfactually—these models were correctly specified to control for risk, there should be no racial/ethnic return difference. (And perhaps, given that these risks are idiosyncratic, efficiency should eliminate the return differences even without risk adjustment.) However, we observe highly significant White–non-White differences of roughly 12 percentage points in Columns (1) through (3).

In Columns (4)–(8), we allow the White–non-White gap to differ between favorites and nonfavorites, as suggested by Figure 4. Columns (5) and (6) include fixed effects to account for the possibility that different trainers might enter horseraces or hire jockeys with systematically different average returns. Among favorites, the White–non-White return gap is 3–5 percentage points, and significant at the 5% level unless horserace fixed effects are included. In particular, results are not driven by differences in the jockeys riding White- and non-White-trained horses.<sup>23</sup> Estimated interaction coefficients in these specifications show that the racial/ethnic return gaps are 15–17 percentage points larger among nonfavorites (significant at the 1% level).

These results suggest that horse bettors discriminate against non-White-named trainers, and that this discrimination is concentrated among nonfavorites. Given the existence of the FLB, nonfavorites offer lower expected returns and may therefore disproportionately attract less-sophisticated or less-informed bettors; it is precisely these bettors who seem to engage in more discrimination.

One concern about interpreting these results as evidence for bias that varies systematically with sophistication or information is the possibility that they are driven by favorites and longshots having different types of trainers rather than different types of bettors. That is, there could be systematic differences in the types of horses trained by White and non-White-named trainers. Estimates including trainer fixed effects, shown in Table 4 Columns (7) and (8), suggest that this concern does not drive our results. These specifications allow the White–non-White return gap to vary with odds, but are based on comparisons between the *same trainers*' horses when they race as favorites versus longshots, thus controlling for all time-invariant signals of trainer quality.<sup>24</sup> The estimated interaction coefficients show that the racial/ethnic return gap remains roughly 15–18 percentage points larger among nonfavorites (significant at the 5% level or higher) using these within-trainer comparisons. This effect is necessarily driven by differences between bets on favorites and longshots rather than differences between favorite and longshot horses' trainers.

The results in Table 4 provide strong evidence that trainers with non-White surnames (measured using 20% population cutoffs) generate higher average returns especially when training longshots (measured using median odds cutoffs). However, there is no reason to expect that these effects should only occur discontinuously across these levels. We ensure the robustness of the results across varying definitions of these two measures.

The downward slope of the FLB means that if less-sophisticated bettors discriminate more, racial/ethnic return differences should vary across the odds distribution. Table 5 shows that this result obtains across a variety of odds splits: pooled 25th percentile odds (only the most-favored horses); within-horserace median

TABLE 5 | Alternative measures of favorite versus nonfavorite.

|                         | Dependent variable: Realized return (\$) |                        |                           |                        |                                |                        |                        |                        |
|-------------------------|--|------------------------|---------------------------|------------------------|--------------------------------|------------------------|------------------------|------------------------|
|                         | (1)                                      | (2)                    | (3)                       | (4)                    | (5)                            | (6)                    | (7)                    | (8)                    |
|                         | p25 Split                                |                        | p50 Split (w/i horserace) |                        | "In the Money" (w/i horserace) |                        | Quintile split         |                        |
| Constant                | -0.1765***<br>(0.0198)                   | -0.1368***<br>(0.0179) | -0.1441***<br>(0.0246)    | -0.1140***<br>(0.0226) | -0.1527***<br>(0.0238)         | -0.1180***<br>(0.0218) | -0.1815***<br>(0.0219) | -0.1414***<br>(0.0206) |
| Non-White               | 0.1214***<br>(0.0284)                    | 0.0082<br>(0.0211)     | 0.1191***<br>(0.0285)     | 0.0371*<br>(0.0221)    | 0.1200***<br>(0.0284)          | 0.0229<br>(0.0215)     | 0.1222***<br>(0.0283)  | 0.0146<br>(0.0221)     |
| Odds (log)              | -0.0826***<br>(0.0203)                   | -0.0843***<br>(0.0202) | -0.0714***<br>(0.0168)    | -0.0738***<br>(0.0167) | -0.0747***<br>(0.0194)         | -0.0770***<br>(0.0194) | -0.0725**<br>(0.0339)  | -0.0791**<br>(0.0337)  |
| Nonfavorite             | 0.1298***<br>(0.0366)                    | 0.0815**<br>(0.0400)   | 0.0770**<br>(0.0307)      | 0.0288<br>(0.0356)     | 0.0882***<br>(0.0339)          | 0.0411<br>(0.0378)     |                        |                        |
| Non-White × Nonfavorite |  | 0.1519***<br>(0.0428)  |                           | 0.1533***<br>(0.0541)  |                                | 0.1517***<br>(0.0477)  |                        |                        |
| Non-White × Odds q2     |  |                        |                           |                        |                                |                        |                        | 0.0141<br>(0.0415)     |
| Non-White × Odds q3     |  |                        |                           |                        |                                |                        |                        | 0.1397**<br>(0.0543)   |
| Non-White × Odds q4     |  |                        |                           |                        |                                |                        |                        | 0.1410**<br>(0.0718)   |
| Non-White × Odds q5     |  |                        |                           |                        |                                |                        |                        | 0.2407**<br>(0.1055)   |
| Odds quintile FE        |  |                        |                           |                        |                                |                        | ✓                      | ✓                      |
| N                       | 74,988                                   | 74,988                 | 74,988                    | 74,988                 | 74,988                         | 74,988                 | 74,988                 | 74,988                 |

Note: The table presents results from estimation of variations of equation 3 using different ways of determining "favorites" from horses' odds. Columns (1) and (2) use the 25th percentile of the pooled odds distribution. Columns (3) and (4) use the within-horserace median (i.e., half the horses in each horserace are considered favorites). Columns (5) and (6) treat as favorites the top three odds-implied finishers in each horserace. Columns (7) and (8) assess separate effects for each odds quintile relative to the first quintile. The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

odds (horses expected to finish in the top half); within-horserace bottom three odds (horses expected to finish "in the money"); and at each quintile of the pooled odds distributions, across which non-White interaction terms vary monotonically as expected.

Our baseline racial/ethnic categorization is based on distinguishing trainer names that bettors may most confidently perceive as White (because they are common, and have low Black and Hispanic population prevalence). We consider robustness across several different ways of assessing trainers' perceived race/ethnicity as a discrete or continuous function of surnames population demographics. Table A.2 breaks non-White-named trainers out into Black and Hispanic (and ambiguous), confirming that the pattern of results is consistent across the groups.<sup>25</sup> In Table 6 Panel (a), we move the Census surname cutoffs to 80% to create a binary categorization that requires names to be treated as much more distinctively Black or Hispanic before being coded as non-White. In Panel (b), we restrict the sample to trainers whose names appear in the Census, and consider return differences

across continuous variation in non-White surname percentage. The results are similar to Table 4 in both panels. Taken together, robustness across approaches suggests our results are driven by racial/ethnic perceptions of trainers' surnames rather than the specific way these surnames are classified.

## 5 | Heterogeneous Effects and Economic Mechanisms

Sophistication varies enormously across bettors (Roeder, 2023), and we complement our earlier results using variation in sophistication associated with the FLB with a number of additional tests assessing where discrimination's effects appear in prices in order to better understand the potential economic mechanisms underlying bettor bias. To do so, we investigate heterogeneous effects of trainer race/ethnicity on realized returns (both overall and especially among longshots) across a number of bet, horse, horserace, and track characteristics.

TABLE 6 | Alternative measures of non-White trainer surnames.

| <b>(a) Binary non-White categorization using 80% Thresholds</b> |                        |                        |                        |                        |                        |                        |                        |                        |
|---|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|   | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Constant  | -0.2365***<br>(0.0136) | -0.1326***<br>(0.0216) | -0.1224***<br>(0.0251) | -0.1010***<br>(0.0242) | -0.0811***<br>(0.0261) | -0.1380***<br>(0.0264) | -0.1553***<br>(0.0259) | -0.2194***<br>(0.0323) |
| Non-White (80%)   | 0.0937**<br>(0.0368)   | 0.0987***<br>(0.0367)  | 0.0988***<br>(0.0367)  | -0.0060<br>(0.0254)    | -0.0426<br>(0.0375)    | 0.0074<br>(0.0283)     | —                      | —                      |
| Odds (log)  |                        | -0.0478***<br>(0.0125) | -0.0625***<br>(0.0194) | -0.0650***<br>(0.0193) | -0.0749***<br>(0.0197) | -0.0470**<br>(0.0199)  | -0.0330*<br>(0.0200)   | -0.0059<br>(0.0220)    |
| Nonfavorite   |                        |                        | 0.0444<br>(0.0376)     | 0.0116<br>(0.0395)     | 0.0199<br>(0.0437)     | 0.0061<br>(0.0404)     | -0.0166<br>(0.0408)    | -0.0067<br>(0.0460)    |
| Non-White (80%) ×<br>Nonfavorite                                |                        |                        |                        | 0.2034***<br>(0.0715)  | 0.2496***<br>(0.0782)  | 0.1942***<br>(0.0723)  | 0.2192***<br>(0.0768)  | 0.2438***<br>(0.0854)  |
| Horsrace FE   |                        |                        |                        |                        | ✓                      |                        |                        | ✓                      |
| Jockey FE   |                        |                        |                        |                        |                        | ✓                      |                        | ✓                      |
| Trainer FE  |                        |                        |                        |                        |                        |                        | ✓                      | ✓                      |
| <i>N</i>  | 74,988                 | 74,988                 | 74,988                 | 74,988                 | 74,988                 | 74,689                 | 74,156                 | 73,856                 |

| <b>(b) Continuous (standardized) non-White measure</b> |                        |                        |                        |                        |                        |                        |                        |                        |
|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|  | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Constant   | -0.2272***<br>(0.0135) | -0.1012***<br>(0.0213) | -0.0932***<br>(0.0255) | -0.0933***<br>(0.0255) | -0.0722***<br>(0.0280) | -0.1395***<br>(0.0279) | -0.1547***<br>(0.0283) | -0.2205***<br>(0.0360) |
| Non-White (z)  | 0.0500***<br>(0.0161)  | 0.0565***<br>(0.0159)  | 0.0567***<br>(0.0159)  | 0.0235**<br>(0.0117)   | 0.0184<br>(0.0175)     | 0.0126<br>(0.0136)     | —                      | —                      |
| Odds (log)   |                        | -0.0571***<br>(0.0132) | -0.0684***<br>(0.0206) | -0.0702***<br>(0.0206) | -0.0839***<br>(0.0214) | -0.0469**<br>(0.0215)  | -0.0351<br>(0.0216)    | -0.0080<br>(0.0244)    |
| Nonfavorite  |                        |                        | 0.0337<br>(0.0401)     | 0.0378<br>(0.0401)     | 0.0560<br>(0.0452)     | 0.0295<br>(0.0411)     | 0.0131<br>(0.0418)     | 0.0296<br>(0.0483)     |
| Non-White (z) ×<br>Nonfavorite                         |                        |                        |                        | 0.0568**<br>(0.0285)   | 0.0680**<br>(0.0315)   | 0.0691**<br>(0.0292)   | 0.0835**<br>(0.0324)   | 0.1123***<br>(0.0368)  |
| Horsrace FE  |                        |                        |                        |                        | ✓                      |                        |                        | ✓                      |
| Jockey FE  |                        |                        |                        |                        |                        | ✓                      |                        | ✓                      |
| Trainer FE   |                        |                        |                        |                        |                        |                        | ✓                      | ✓                      |
| <i>N</i>   | 66,179                 | 66,179                 | 66,179                 | 66,179                 | 66,155                 | 65,882                 | 65,461                 | 65,138                 |

Note: The table presents results from estimation of variations of Equation (3) using different measures of surname classification. In Panel (a), nonwhite surnames are defined using a discrete 80% Hispanic/Black prevalence threshold in the Census (rather than the 20% used in the main analysis); using this measure, 19% of trainers are classified as having non-White surnames. In Panel (b), non-White is a standardized, continuous measure of the nonwhite fraction of individuals in the Census with a given surname. Sample size is slightly reduced using the continuous measure since we drop observations for trainers whose surnames are not in the Census file. The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1. \*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

First, we examine heterogeneous effects across different types of bets on the same horse. A win, place, or show bet pays out if the picked horse finishes first, top two, or top three, respectively. Whereas less-sophisticated bettors may typically focus on win bets, serious handicappers' more sophisticated, hedged betting

strategies typically recommend the use of place and show bets (together with exotic wagers) (Cronley, 2000; Nelson, 2020). Variation in racial/ethnic return gaps across bet types therefore help understand the relationship between bettor bias and bettor sophistication.

**TABLE 7** | Win, place, and show realized returns by odds and White/non-White trainer surname.

|                         | <b>Dependent Variable: Realized Return (\$)</b> |                            |                           |                          |                            |                           |
|-------------------------|---|----------------------------|---------------------------|--------------------------|----------------------------|---------------------------|
|                         | <b>(1)</b><br><b>Win</b>                        | <b>(2)</b><br><b>Place</b> | <b>(3)</b><br><b>Show</b> | <b>(4)</b><br><b>Win</b> | <b>(5)</b><br><b>Place</b> | <b>(6)</b><br><b>Show</b> |
| Constant                | -0.1880***<br>(0.0449)                          | -0.0919***<br>(0.0235)     | -0.0627***<br>(0.0147)    | -0.1326***<br>(0.0405)   | -0.0805***<br>(0.0229)     | -0.0620***<br>(0.0147)    |
| Non-White               | 0.1671***<br>(0.0482)                           | 0.0416<br>(0.0258)         | 0.0035<br>(0.0167)        | 0.0239<br>(0.0337)       | 0.0120<br>(0.0206)         | 0.0016<br>(0.0154)        |
| Odds (log)              | -0.0417<br>(0.0318)                             | -0.0815***<br>(0.0174)     | -0.0947***<br>(0.0107)    | -0.0468<br>(0.0315)      | -0.0825***<br>(0.0174)     | -0.0947***<br>(0.0108)    |
| Nonfavorite             | 0.0373<br>(0.0614)                              | 0.0532<br>(0.0354)         | 0.0365<br>(0.0248)        | -0.0550<br>(0.0699)      | 0.0341<br>(0.0391)         | 0.0352<br>(0.0271)        |
| Non-White × Nonfavorite |   |                            |                           | 0.2791***<br>(0.0945)    | 0.0578<br>(0.0507)         | 0.0037<br>(0.0329)        |
| <i>N</i>                | 28,390  | 28,390                     | 28,390                    | 28,390                   | 28,390                     | 28,390                    |

Notes: The table presents results from estimation of variations of Equation (3). The dependent variable in Columns (1) and (4) is the realized return from a \$1 win bet (i.e., the horse bet on comes in first); these results replicate analysis from Columns (3) and (4) of Table 4 using the subsample for which place and show payouts are available. The dependent variable in Columns (2)–(3) and (5)–(6) are the realized returns from \$1 place or show bets (i.e., the horse bet on finishes in the top two, or top three). Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from the Racing Post–Horse Racing Nation matched sample and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

Results for win, place, and show bets are in Table 7. We find that our results for win bets hold in the matched Racing Post–Horse Racing Nation sample. Specifically, Columns (1) and (4) show that win bets on non-White-named trainers earn 17 percentage point higher average returns than bets on White-named trainer, an effect driven by longshots, where the return gap is 28 percentage points larger than for favorites. Columns (2)–(3) and (5)–(6) show that as we move to the place and show pools, the estimated effects of discrimination decrease monotonically in magnitude, and are only statistically significant for win bets. Racial/ethnic gaps, both in average realized returns and in the strength of the FLB, may exist only for win bets, suggesting that discrimination is strongest among the least sophisticated bettors.

Next, we examine how horses' previous performance (momentum) plays a role in determining discrimination's effects on financial returns. The way a horse ran in its prior horseraces is informative about its current prospects; that is, operating performance shows persistence. Sophisticated bettors may therefore be disproportionately likely to bet momentum strategies, inferring quality from fundamental signals rather than trainer surnames. Less-sophisticated bettors should be over-represented among horses with poor prior performance.

In Panel (a) of Table 8, we present estimates of Equation (3) (as in Table 4, Columns 3–4), partitioning the sample between horses that finished in the bottom versus the top half of their previous horserace. Figures 1 and IA.1 illustrate how information about previous finishing positions is available to bettors. The results suggest that bettors on horses with poor prior performance are the most susceptible to racial/ethnic discrimination, especially those betting on nonfavorites. Comparing the estimates in Columns (1)–

(3) shows that among horses with a previous finish in the bottom half, non-White-named trainers generate 17 percentage point higher returns, almost triple the (statistically insignificant) 6 percentage point gap for top-half finishers. Column (2) shows that nonfavorites drive this difference. Whereas racial/ethnic differences in the strength of the FLB are large among poor prior performers, we find a much smaller effect among the horses that ran well in their previous race and therefore attract more-informed bettors.

Next, we consider whether the effects of discrimination vary across purse size. The purse represents prize money paid to horses' owners, trainers, and jockeys (based on finishing position), and the best and best-known horses typically race for larger purses. If the pricing effects of discrimination are driven by less-informed or less-sophisticated bettors, these effects may be stronger in small-purse horseraces for at least two reasons. First, small purses may be associated with reduced motivation for both bettors and tracks to gather comprehensive information on trainer, jockey, and horse quality, and acquiring information on lower profile horserace participants may be more challenging. (This is related to the argument and findings in Gandhi and Serrano-Padial, 2015.) Second, if mispricing does exist, smaller-purse horseraces may be less attractive to arbitrageurs seeking to profit off market inefficiencies. Lower liquidity in betting markets for smaller-purse horseraces may increase the likelihood of observing pricing effects of discrimination against non-White-named trainers.

Panel (b) of Table 8 shows results split by purse size. Comparing Columns (1) and (3), we find the average White–non-White return gap is slightly larger in small-purse horseraces (14 vs. 10

TABLE 8 | Heterogeneous effects across horse/horserace characteristics.

| <b>(a) Prior finishing position</b> |                                 |                        |                               |                        |
|-------------------------------------|---------------------------------|------------------------|-------------------------------|------------------------|
|                                     | <b>(1)</b>                      | <b>(2)</b>             | <b>(3)</b>                    | <b>(4)</b>             |
|                                     | <b>Prior finish bottom half</b> |                        | <b>Prior finish top half</b>  |                        |
| Constant                            | -0.1359**<br>(0.0672)           | -0.0840<br>(0.0606)    | -0.2038***<br>(0.0382)        | -0.1943***<br>(0.0371) |
| Non-White                           | 0.1650***<br>(0.0622)           | 0.0274<br>(0.0460)     | 0.0636<br>(0.0410)            | 0.0377<br>(0.0293)     |
| Odds (log)                          | -0.0545<br>(0.0429)             | -0.0594<br>(0.0424)    | -0.0091<br>(0.0322)           | -0.0098<br>(0.0323)    |
| Nonfavorite                         | 0.0583<br>(0.0771)              | -0.0118<br>(0.0874)    | -0.0136<br>(0.0576)           | -0.0363<br>(0.0654)    |
| Non-White × Nonfavorite             |                                 | 0.2358**<br>(0.1110)   |                               | 0.0729<br>(0.1067)     |
| <i>N</i>                            | 20,715                          | 20,715                 | 26,450                        | 26,450                 |
| <b>(b) Purse size</b>               |                                 |                        |                               |                        |
|                                     | <b>(1)</b>                      | <b>(2)</b>             | <b>(3)</b>                    | <b>(4)</b>             |
|                                     | <b>Small purse</b>              |                        | <b>Large purse</b>            |                        |
| Constant                            | -0.1192***<br>(0.0319)          | -0.0804***<br>(0.0299) | -0.1684***<br>(0.0436)        | -0.1483***<br>(0.0396) |
| Non-White                           | 0.1407***<br>(0.0374)           | 0.0398<br>(0.0285)     | 0.1010**<br>(0.0435)          | 0.0475<br>(0.0304)     |
| Odds (log)                          | -0.0941***<br>(0.0231)          | -0.0975***<br>(0.0230) | -0.0340<br>(0.0308)           | -0.0362<br>(0.0305)    |
| Non-favorite                        | 0.0928*<br>(0.0504)             | 0.0248<br>(0.0558)     | -0.0006<br>(0.0556)           | -0.0300<br>(0.0626)    |
| Non-White × Nonfavorite             |                                 | 0.2030***<br>(0.0750)  |                               | 0.1060<br>(0.0855)     |
| <i>N</i>                            | 37,516                          | 37,516                 | 37,472                        | 37,472                 |
| <b>(c) Track conditions</b>         |                                 |                        |                               |                        |
|                                     | <b>(1)</b>                      | <b>(2)</b>             | <b>(3)</b>                    | <b>(4)</b>             |
|                                     | <b>Fast track conditions</b>    |                        | <b>Other track conditions</b> |                        |
| Constant                            | -0.1697***<br>(0.0368)          | -0.1320***<br>(0.0337) | -0.1062***<br>(0.0387)        | -0.0877**<br>(0.0358)  |
| Non-White                           | 0.1142***<br>(0.0401)           | 0.0159<br>(0.0274)     | 0.1244***<br>(0.0402)         | 0.0750**<br>(0.0317)   |
| Odds (log)                          | -0.0520*<br>(0.0274)            | -0.0554**<br>(0.0273)  | -0.0812***<br>(0.0266)        | -0.0833***<br>(0.0264) |
| Nonfavorite                         | 0.0669<br>(0.0530)              | 0.0031<br>(0.0597)     | 0.0316<br>(0.0528)            | 0.0036<br>(0.0586)     |
| Non-White × Nonfavorite             |                                 | 0.1990**<br>(0.0807)   |                               | 0.0964<br>(0.0790)     |
| <i>N</i>                            | 39,254                          | 39,254                 | 35,734                        | 35,734                 |

(Continues)

**TABLE 8** | (Continued)

Note: The table presents results from estimation of variations of Equation (3). The dependent variable is the realized return from a \$1 win bet. Panel (a) splits the sample by prior finishing position (momentum). Panel (b) splits the sample by the average purse size. Panel (c) splits the sample by track conditions. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

percentage points). Results in Columns (2) and (4) show that once again, pricing effects of discrimination are driven by nonfavorites, with a highly significant effect (of nearly double the magnitude) for the small-purse horseraces where we expect less-sophisticated bettors to be more prevalent, and arbitrage to be less effective in eliminating mispricing.

We complement these momentum- and size-based heterogeneous effects tests with an additional horserace-level analysis using variation in track conditions. Horses perform differently based on the quality of dirt and turf at the track, which vary with recent weather conditions. On a “fast” dirt track—the most common condition—unsophisticated bettors may on average represent a larger fraction of the gambling pool. In contrast, conditions like “muddy” and “sloppy” are associated with bad weather that may lead casual bettors to stay home. Dirt track conditions other than “fast,” as well as racing on turf, also reward higher handicapping skill since horse performance is more difficult to predict.

In Panel (c) of Table 8, we show estimates splitting between “fast” and other track conditions. While the average White–non-White return gap is an indistinguishable 11–12 percentage points on both types of track conditions (Columns 1 and 3), we find that for fast conditions, this effect is entirely driven by longshots. These horses and these conditions—precisely where we expect the least informed and least sophisticated bettors to have the strongest pricing effects—are where we see evidence of the strongest racial/ethnic discrimination. Interestingly, with nonfast conditions, we find statistically significant racial/ethnic pricing gaps among favorites, although the differences between favorites and longshots are smaller than for fast tracks (and statistically insignificant).

## 6 | Conclusion

We examine the impact of racial/ethnic bias on pricing in horse betting markets, characterized by significant limits to arbitrage, short termination windows, posted payouts, extensive public information about fundamentals, and an absence of systematic risk. We consider return differences across horses whose trainers have racially/ethnically distinctive surnames, which bettors may see as a proxy for horse quality or a source of nonpecuniary returns. The results show that horses with non-White-named trainers earn higher realized returns, finish further ahead in the field, and win more often than would be predicted by their odds, consistent with bettor discrimination against these trainers. Our findings thus show that discrimination, whether rooted in animus (taste-based) or inaccurate beliefs about quality (statistical), imposes a direct financial cost on biased bettors—a result consistent with models where bias is costly for the agent who holds it (Becker, 1971).

Risk-adjusting to control for the well-known “favorite–longshot” bias, the study finds that non-White-named trainers earn higher returns especially among longshots. The racial/ethnic return differences we document are stronger—overall and especially among longshots—for “win” than “place” and “show” bets, among horses with poor prior performance, and in low-stakes races with “fast” conditions. These results are consistent with the effects of discrimination being strongest among less-informed and less-sophisticated bettors.

We contribute to literatures on horse race gambling markets, and on fair pricing and racial bias in asset pricing more generally, elucidating the impact of ethnicity on investor behavior in financial markets. The findings provide empirical evidence of racial biases in horse betting, highlighting the presence of pricing distortions due to discrimination. While comparisons of realized returns have long been used to analyze betting biases including the FLB, we believe we are the first to apply this approach to discrimination in horseracing. We show how the financial effects of discrimination can vary even within a market, considering sources of cross-sectional and time-series variation to show that the pricing effects of bias are driven by the least sophisticated or informed bettors, and market segments where limits to arbitrage prevent these effects from getting competed away. Although our specific tests may be uniquely applicable to horse racing, they suggest that discrimination may have the largest effects where limits to arbitrage are strong and investors are unsophisticated or poorly informed, hypotheses worth investigating using related approaches in other markets.

## Endnotes

<sup>1</sup>Moskowitz (2021) shows how bets on leagues such as the NFL, NHL, NBA, and MLB have these two critical distinctions from traditional financial markets, which limit the role of confounds that can complicate inference in other financial settings. Analysis of betting markets can generate insight on market equilibrium (Quandt, 1986), insider trading (Schnytzer and Shilony, 1995), investing factors (Moskowitz, 2021), preferences (Moskowitz and Vasudevan, 2022), and market efficiency (surveyed in Vandenbruaene et al., 2022).

<sup>2</sup>When betting is *not* parimutuel, winning bettors typically receive different payouts that depend on where and when their bets were placed. Aside from making research more difficult, this has implications for investor updating, suggesting that learning about mispricing might happen more quickly in parimutuel settings. Larsen et al. (2008) and Igan et al. (2015) show that returns on book-made NBA bets vary with teams’ racial composition, although referee bias (as documented in Price and Wolfers, 2010) may play a role unlikely to exist in the horseracing context. Researchers have a long history of attempting to answer economic questions from the horse track (see, e.g., Snyder, 1978). Previous work extrapolates evidence from horse betting markets to answer broader questions about financial markets, assessing market efficiency (Figlewski, 1979; Hausch et al., 1981), informed behavior

(Asch et al., 1982), insider knowledge (Crafts, 1985; Law and Peel, 2002), anomalies with lottery-like preferences (Thaler and Ziemba, 1988; Golec and Tamarkin, 1998; Snowberg and Wolfers, 2010), arbitrage (Hausch and Ziemba, 1990), market manipulation (Camerer, 1998), market timing (Forrest et al., 2018), and market maker profit (Green et al., 2020).

<sup>3</sup>Limited attention likely also plays a role, given the large number of trainers active in the United States, and the large number of horses racing on a given day. Bettors may lack the inclination or ability to assess quality directly, and therefore rely on heuristics.

<sup>4</sup>While surnames are highly salient, bettors are not typically exposed to other signals of trainer demographics except for a very small number of celebrity trainers. Although bettors might be reluctant to admit to reliance on racial/ethnic stereotypes (and may not even know they are subject to implicit bias), debate over whether an attribute is correlated with quality is perhaps the most common discussion topic at the track. A strategy that some praise as handicapping can be derided as mere superstition by others, echoing attempts to differentiate “accurate” from “inaccurate statistical discrimination” in other economic contexts.

<sup>5</sup>Quick resolution of uncertainty in our context means that both market participants and researchers can learn about the payoff-relevance of various quality signals more easily than when assessing bias in financial intermediation or related settings (as in, e.g., Bayer et al., 2018; Dougal et al., 2019; Fuster et al., 2022; Butler et al., 2023). We may also face less concern than in these settings that race/ethnicity is correlated with omitted variables that should drive investor behavior.

<sup>6</sup>We control for any differences in the jockeys hired by White and non-White-named trainers using jockey fixed effects, and find that they do not substantively affect our main results. We also explicitly consider jockey race/ethnicity in Table A.4. Unlike trainers—who play a long-term strategic role in determining horse performance—U.S. jockeys are overwhelmingly non-White (approximately 72% using our baseline methodology in our sample) and we do not find evidence for racial/ethnic discrimination here where bettors may be more used to non-White individuals playing a more short-term, tactical role. While anecdotal evidence suggests that superstitious bettors may pay attention to horse names, these are not typically racialized.

<sup>7</sup>None of these papers look explicitly at financial returns, focusing instead on operating outcomes (e.g., win probabilities and horse finish positions). While the U.S. context is characterized by parimutuel betting, Brown and Yang (2015) and Cashmore et al. (2022) consider fixed-odds bets in the United Kingdom where implied win probabilities are calculated using prices from various bookmakers.

<sup>8</sup>Race and ethnicity affect investment in a number of financial assets with potentially unsophisticated investors and limits to arbitrage, including bonds (Dougal et al., 2019), mutual funds (Kumar et al., 2015; Niessen-Ruenzi and Ruenzi, 2019; Han et al., 2022), hedge funds (Aggarwal and Boyson, 2016; Lu et al., 2024), entrepreneurial ventures (Ewens and Townsend, 2020; Hebert, 2023; Gornall and Strebulaev, 2024), peer-to-peer loans (Ravina, 2019; D’Acunto et al., forthcoming; Diep-Nguyen et al., 2025), and art and collectibles (Adams et al., 2021; Kim and Lee, 2022).

<sup>9</sup>Our discussion and analysis focus primarily on bets on a single horse to win a race, but many features are similar for “place” and “show” wagers (which we consider in Table 7), as well as more exotic bets.

<sup>10</sup>Analyzing trainers is a cornerstone of sophisticated betting strategy, and the handicapping literature is filled with discussions of statistical patterns based on a trainer’s past performance in specific situations. Trainers’ importance comes through in advice to bettors, and “there is certainly a general feeling that the skill of the trainer is the most important thing when it comes to the possible success of the horse” (UnderstandingRacing.com, “Jockey vs. Trainer—Who Is The Most Important?”).

<sup>11</sup>We follow the AP Stylebook in capitalizing “Black” and “Hispanic” but not “White.” Table IA.1 lists the sample’s most racially/ethnically

distinctive trainer surnames. “Denelsbeck” is a predominately White surname, with 99% of the U.S. population with that surname classified as White in the Census. Analogously, “Velazquez” is an overwhelmingly Hispanic surname (96% Hispanic) and Ivory is a highly Black surname (73% Black). The most White- and Hispanic-named trainers’ perceived race/ethnicities are much more certain than those of Black-named trainers, with percentages closer to 100%. In contrast, even the 10 most distinctively Black-named trainers have surname percentages averaging only around 75% Black.

<sup>12</sup>Summary statistics are reported over the sample of horses  $\times$  horseraces, so trainers are weighted by the total number of times their horses race, and horseraces are weighted by their number of runners.

<sup>13</sup>The distributions of odds-predicted and actual finishing positions are slightly different, given the presence of horses in the same race with equal starting odds and the fact that horses can finish in a tie.

<sup>14</sup>This corresponds to the  $-22\%$  loss suffered by a bettor who held an equal-weighted portfolio across every horse in every race and is close to the average realized returns of  $-\$0.23$  in Snowberg and Wolfers (2010) and  $-\$0.25$  in Green et al. (2020). The negative mean is driven by the track’s “takeout” from each parimutuel pool (although the aggregate takeout corresponds to the return on a *value-weighted* portfolio and is typically less than 22%). Takeouts vary somewhat over time and across tracks; our slightly less negative average return than other studies’ may also relate to the Racing Post sample’s heavy weight on stakes races.

<sup>15</sup>Citetottaviani2008favorite provide a useful review of many proposed explanations for the FLB. Neoclassical interpretations of the FLB may center gamblers who place excessive bets on longshots because of a love of risk (Weitzman, 1965; Ali, 1977; Quandt, 1986; Golec and Tamarkin, 1998). In contrast, behavioral theories may suggest misperceptions of win probabilities (Griffith, 1949; Snowberg and Wolfers, 2010; Hundtofte and Meyer, 2023), belief heterogeneity (Gandhi and Serrano-Padial, 2015), or heterogeneous, unexpected-utility preferences (Chiappori et al., 2019). Even more simply, uninformed bettors may wager randomly (Hurley and McDonough, 1995). Green et al. (2020) argue that the FLB is related to the suggestive “morning line” odds published by tracks’ professional handicappers, which serve to attract bettors and maximize track profit.

<sup>16</sup>We do not find statistically significant evidence for a level break at median odds, which corresponds to the *Nonfavorite* coefficient in Column (3). In Columns (4) and (5), this coefficient represents differences in the intercept, rather than the size of a level break (if there were one).

<sup>17</sup>Proposed explanations of the FLB include bettors with risk-seeking preferences (or perhaps a “recreational interest” in gambling, as in Ottaviani and Sørensen, 2010) or heterogeneous beliefs about horses’ prospects (as in Gandhi and Serrano-Padial, 2015). Several papers also argue that the FLB is empirically stronger in lower-information environments (e.g., Gandar et al., 2001; Gandhi and Serrano-Padial, 2015; Ziemba, 2019).

<sup>18</sup>Mean finishing position for White versus non-White: 4.90 versus 4.91, *t*-test for unequal means  $p = 0.54$ . Standard deviation: 2.88 versus 2.88, *F*-test for unequal variances  $p = 0.75$ . Wilcoxon rank-sum (Mann–Whitney) test:  $p = 0.32$ . There are however significant differences in finishing position *when controlling for odds* as shown in Table 3 Columns (3)–(4) and discussed below, consistent with racial/ethnic differences in bettor behavior rather than operating performance.

<sup>19</sup>These average returns are illustrated in Figure IA.2. The white–non-White difference is significant at the 1% level (Table 4, Column 1), while each of the three non-White-named groups offer average returns that exceed the White-named trainers’ by amounts significant at the 5% and 1% levels (Table A.2, Column 1).

<sup>20</sup>Figure IA.3 plots the odds distribution separately for White- and non-White-named trainers. Non-White-named trainers are also somewhat more likely than White-named trainers to train the most extreme favorites (offering lower than 1–1 odds). However, the difference is small

relative to those that exist at the longshot end of the odds distribution, and the FLB is relatively flat (i.e., weak) among these extreme favorites.

<sup>21</sup>Tracks employ professional handicappers who publish “morning line” (M/L) odds that attempt to predict (and perhaps shape, as in Green et al., 2020) bettor behavior. They are not bookmakers, and while their “odds” have no direct relevance to payouts, they are highly salient. To understand whether the morning line also reflects racial/ethnic mispricing, we consider a subsample for which M/L odds data are available. In Table A.3, we repeat the analysis of Table 3 using our Racing Post–Horse Racing Nation matched sample and replace measures calculated using starting parimutuel odds with analogs using M/L odds. That is, we analyze counterfactual returns that would have been earned if starting odds were equal to those set by the track’s morning line handicapper. We also consider heterogeneity in how morning line odds predict where horses finish and whether they win; results using starting and M/L odds generate similar racial/ethnic outcome differences. (We find in untabulated regressions that the differences between decimal M/L and final odds do not systematically vary by race/ethnicity, following Igan et al., 2015, who assess changes in basketball point spreads as a function of teams’ racial composition.) Our results suggest that track handicappers may be driving and/or (roughly accurately) predicting the average level of bettor discrimination, although we cannot distinguish between these two channels. We appreciate Bruce Carlin for suggesting this line of analysis.

<sup>22</sup>To account for the possibility that unobserved factors at the horserace, trainer, or jockey level could lead to correlated error terms, we test the robustness of our results by clustering standard errors at each of these levels. We find (in untabulated results) that the resulting standard errors are nearly unchanged, suggesting that any within-cluster correlation of the errors is either weak or, more importantly, not systematically related to the covariates in our models, and therefore does not substantively affect our inferences.

<sup>23</sup>We also replicate in Table A.4 the full set of specifications from Table 4 using jockey surnames rather than trainer surnames. We find no statistically nor economically significant differences in the returns earned on non-White-named and White-named jockeys, perhaps because the latter are rare.

<sup>24</sup>We also consider controls for several time-varying observable trainer characteristics in Table A.5. Column (1) includes Trainer Won Last Race, which is an indicator variable that takes the value of 1 if a trainer has a horse that won their previous race to control for the perceived “hot streak” of the trainer. Column (2) includes Number of Races in a Year for a Trainer (so far) to capture “Baffert effects” where well-known and popular trainers enter horses in more races. Column (3) includes Trainer Win Percentage for a Track–Year (so far) to control for home-track effects. Column (4) includes all the observable controls together with trainer fixed effects. Across columns, the estimated interaction coefficients show that the racial/ethnic gap remains between 15 and 23 percentage points larger among nonfavorites, with differences significant at the 1% level.

<sup>25</sup>Figure IA.4 plots realized returns as a function of log fractional odds separately for White-, Hispanic-, and Black-named trainers. As in Figure 4, realized return averages for each odds quintile of Hispanic- and Black-named trainers are greater than White-named ones (except for 5-1 Hispanic-named). However, there is a more limited range of odds at which the nonparametric fits’ confidence intervals are nonoverlapping: Roughly 5–1 to 20–1 odds for Hispanic- and Black- versus White-named trainers. At no point in the odds range are Hispanic- and Black-named trainers’ realized returns statistically different from each other.

## References

Adams, R. B., R. Kräussl, M. Navone, and P. Verwijmeren. 2021. “Gendered Prices.” *Review of Financial Studies* 34, no. 8: 3789–3839.

Aggarwal, R., and N. M. Boyson. 2016. “The Performance of Female Hedge Fund Managers.” *Review of Financial Economics* 29: 23–36.

Ali, M. 1977. “Probability and Utility Estimates for Racetrack Bettors.” *Journal of Political Economy* 85, no. 4: 803–815.

Arai, M., and P. Thoursie. 2009. “Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings.” *Journal of Labor Economics* 27, no. 1: 127–147.

Asch, P., B. Malkiel, and R. Quandt. 1982. “Racetrack Betting and Informed Behavior.” *Journal of Financial Economics* 10, no. 2: 187–194.

Barghouthy, P., S. Corbett-Davies, D. Jenson, et al. 2020. “A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States.” *Nature Human Behaviour* 4, no. 7: 736–745.

Bayer, P., F. Ferreira, and S. Ross. 2018. “What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders.” *Review of Financial Studies* 31, no. 1: 175–205.

Becker, G. S. 1971. *The Economics of Discrimination*. University of Chicago Press.

Bertrand, M., and S. Mullainathan. 2004. “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94, no. 4: 991–1013.

Binder, A., and P. Grimes. 2022. “Defying the Odds: Horseplayers and Female Jockeys.” *Applied Economics Letters* 29, no. 17: 1556–1563.

Binder, A., P. Grimes, and R. Winterbotham. 2021. “In the Money: Gender and Jockey Success on the Thoroughbred Racetrack.” *Journal of Sports Economics* 22, no. 3: 295–328.

Brown, A., and F. Yang. 2015. “Does Society Underestimate Women? Evidence From the Performance of Female Jockeys in Horse Racing.” *Journal of Economic Behavior & Organization* 111: 106–118.

Butler, A., E. Mayer, and J. Weston. 2023. “Racial Disparities in the Auto Loan Market.” *Review of Financial Studies* 36, no. 1: 1–41.

Camerer, C. 1998. “Can Asset Markets Be Manipulated? A Field Experiment With Racetrack Betting.” *Journal of Political Economy* 106, no. 3: 457–482.

Cashmore, V., N. Coster, D. Forrest, I. McHale, and B. Buraimo. 2022. “Female Jockeys – What Are the Odds?” *Journal of Economic Behavior & Organization* 202: 703–713.

Chiappori, P., A. Gandhi, B. Salanié, and F. Salanié. 2019. “From Aggregate Betting Data to Individual Risk Preferences.” *Econometrica* 87, no. 1: 1–36.

Crafts, N. 1985. “Some Evidence of Insider Knowledge in Horse Race Betting in Britain.” *Economica* 52, no. 207: 295–304.

Cronley, J. 2000. *The Not-So-Glorious Life of Show Betting*. ESPN.

D’Acunto, F., P. Ghosh, and A. G. Rossi. 2026. “How Costly Are Cultural Biases? Evidence From FinTech.” *Journal of Financial Economics* 175: 104202.

Diep-Nguyen, H., M. K. Price, and J. Yang. 2025. “Racial and Gender Favoritism in Crowdfunding—Evidence From the Field.” SSRN, 5272180.

Dougal, C., P. Gao, W. Mayew, and C. Parsons. 2019. “What’s in a (School) Name? Racial Discrimination in Higher Education Bond Markets.” *Journal of Financial Economics* 134, no. 3: 570–590.

Eggers, A., R. Freier, V. Grembi, and T. Nannicini. 2018. “Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions.” *American Journal of Political Science* 62, no. 1: 210–229.

Ewens, M., and R. Townsend. 2020. “Are Early Stage Investors Biased Against Women?” *Journal of Financial Economics* 135, no. 3: 653–677.

Feess, E., H. Müller, and C. Schumacher. 2014. “The Favorite–Longshot Bias and the Impact of Experience.” *Business Research* 7: 217–234.

Feigenberg, B., and C. Miller. 2022. “Would Eliminating Racial Disparities in Motor Vehicle Searches Have Efficiency Costs?” *Quarterly Journal of Economics* 137, no. 1: 49–113.

- Fenly, B. 2023. "Purses Set Record in 2022, But Handle Drops Slightly." *Thoroughbred Daily News*.
- Figlewski, S. 1979. "Subjective Information and Market Efficiency in a Betting Market." *Journal of Political Economy* 87, no. 1: 75–88.
- Forrest, D., T. Kainulainen, J. Saastamoinen, and N. Suhonen. 2018. "Is Timing Everything in Horse Betting? Bet Amount, Timing and Bettors' Returns in Pari-Mutuel Wagering Markets." *Economics Letters* 173: 97–99.
- Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *Journal of Finance* 77, no. 1: 5–47.
- Gandar, J. M., R. A. Zuber, and R. S. Johnson. 2001. "Searching for the Favourite-Longshot Bias Down Under: An Examination of the New Zealand Pari-Mutuel Betting Market." *Applied Economics* 33, no. 13: 1621–1629.
- Gandhi, A., and R. Serrano-Padial. 2015. "Does Belief Heterogeneity Explain Asset Prices: The Case of the longshot Bias." *Review of Economic Studies* 82, no. 1: 156–186.
- Geertsema, P., and C. R. Schumacher. 2016. "Parimutuel Contests With Strategic Risk-Sensitive Bettors." *Applied Economics* 48, no. 12: 1140–1158.
- Golec, J., and M. Tamarkin. 1998. "Bettors Love Skewness, Not Risk, at the Horse Track." *Journal of Political Economy* 106, no. 1: 205–225.
- Gornall, W., and I. A. Strebulaev. 2024. "Gender, Race, and Entrepreneurship: A Randomized Field Experiment on Venture Capitalists and Angels." *Management Science* 71: 5308–5327.
- Green, E., H. Lee, and D. Rothschild. 2020. "The Favorite–Longshot Midas," SSRN, 3271248.
- Griffith, R. 1949. "Odds Adjustments by American Horse-Race Bettors." *American Journal of Psychology* 62, no. 2: 290–294.
- Grumbach, J., and A. Sahn. 2020. "Race and Representation in Campaign Finance." *American Political Science Review* 114, no. 1: 206–221.
- Guell, M., J. Rodriguez Mora, and C. Telmer. 2015. "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating." *Review of Economic Studies* 82, no. 2: 693–735.
- Han, L., X. Huang, O. Kadan, and J. Wu. 2022. "The True Colors of Money: Racial Diversity and Asset Management." SSRN, 3966470.
- Hausch, D., M. Rubinstein, and W. Ziemba. 1981. "Efficiency of the Market for Racetrack Betting." *Management Science* 27, no. 12: 1435–1452.
- Hausch, D., and W. Ziemba. 1990. "Arbitrage Strategies for Cross-Track Betting on Major Horse Races." *Journal of Business* 63: 61–78.
- Hebert, C. 2023. "Gender Stereotypes and Entrepreneur Financing." SSRN, 3318245.
- Hundtofte, S., and A. Meyer. 2023. "The Longshot Bias Is a Context Effect." *Management Science* 69, no. 11: 6954–6968.
- Hurley, W., and L. McDonough. 1995. "A Note on the Hayek Nypothesis and the Favorite–Longshot Bias in Parimutuel Betting." *American Economic Review* 85: 949–955.
- Igan, D., M. Pinheiro, and J. Smith. 2015. "A Study of a Market Anomaly: 'White Men Can't Jump', but Would You Bet on It?" *Journal of Economic Behavior & Organization* 113: 13–25.
- Kim, J.-E., and Y. Lee. 2022. "Racial Disparities and Potential Channels in the NBA Digital Collectibles Market." SSRN, 4236789.
- Kumar, A., A. Niessen-Ruenzi, and O. Spalt. 2015. "What's in a Name? Mutual Fund Flows When Managers Have Foreign-Sounding Names." *Review of Financial Studies* 28, no. 8: 2281–2321.
- Larsen, T., J. Price, and J. Wolfers. 2008. "Racial Bias in the NBA: Implications in Betting Markets." *Journal of Quantitative Analysis in Sports* 4, no. 2: 1–19.
- Law, D., and D. Peel. 2002. "Insider Trading, Herding Behaviour and Market Plungers in the British Horse Race Betting Market." *Economica* 69, no. 274: 327–338.
- Leeds, M., and H. Rockoff. 2019. "Beating the odds: Black jockeys in the Kentucky Derby 1870–1911." *Historical Perspectives on Sports Economics* 1: 136–149.
- Leeds, M., and H. Rockoff. 2020. "Jim Crow in the Saddle: The Expulsion of African American Jockeys From American Racing." 28167. National Bureau of Economic Research, 1–58.
- Lu, Y., N. Y. Naik, and M. Teo. 2024. "Race, Discrimination, and Hedge Funds." SSRN, 4070484.
- Moskowitz, T. 2021. "Asset Pricing and Sports Betting." *Journal of Finance* 76, no. 6: 3153–3209.
- Moskowitz, T., and K. Vasudevan. 2022. "Betting Without Beta." SSRN, 3845505.
- Nelson, A. 2020. "What Is a Show Bet in Horse Racing?" Offshore Sports.
- Niessen-Ruenzi, A., and S. Ruenzi. 2019. "Sex Matters: Gender Bias in the Mutual Fund Industry." *Management Science* 65, no. 7: 3001–3025.
- Ottaviani, M., and P. Sørensen. 2008. "The Favorite–Longshot Bias: An Overview of the Main Explanations." In *Handbook of Sports and Lottery Markets*. Vol. 1. 83–101. Elsevier.
- Ottaviani, M., and P. Sørensen. 2010. "Noise, Information, and the Favorite–Longshot Bias in Parimutuel Predictions." *American Economic Journal: Microeconomics* 2, no. 1: 58–85.
- Price, J., and J. Wolfers. 2010. "Racial Discrimination Among NBA Referees." *Quarterly Journal of Economics* 125, no. 4: 1859–1887.
- Quandt, R. 1986. "Betting and Equilibrium." *Quarterly Journal of Economics* 101, no. 1: 201–207.
- Ravina, E. 2019. "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets." SSRN, 1107307.
- Roeder, O. 2023. "I Used AI to Bet on Horse-Racing. Here's What happened." *Financial Times*.
- Schnytzer, A., and Y. Shilony. 1995. "Inside Information in a Betting Market." *Economic Journal* 105, no. 431: 963–971.
- Snowberg, E., and J. Wolfers. 2010. "Explaining the Favorite–Longshot Bias: Is It Risk-Love or Misperceptions?" *Journal of Political Economy* 118, no. 4: 723–746.
- Snyder, W. 1978. "Horse Racing: Testing the Efficient Markets Model." *Journal of Finance* 33, no. 4: 1109–1118.
- Thaler, R., and W. Ziemba. 1988. "Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries." *Journal of Economic Perspectives* 2, no. 2: 161–174.
- Vandenbruaene, J., M. D. Ceuster, and J. Annaert. 2022. "Efficient Spread Betting Markets: A Literature Review." *Journal of Sports Economics* 23, no. 7: 907–949.
- Weitzman, M. 1965. "Utility Analysis and Group Behavior: An Empirical Study." *Journal of Political Economy* 73, no. 1: 18–26.
- Yakowicz, W. 2023. "U.S. Set Gambling Record in 2022 With More Than \$54.9 Billion in Revenue." *Forbes*.
- Ziemba, W. 2019. *Exotic Betting at the Racetrack*. World Scientific.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.  
Data S1

## Appendix

TABLE A.1 | Variable definitions.

| Variable                        | Definition   |
|---------------------------------|--|
| <b>Horserace/Bet outcomes</b>   |  |
| Realized return                 | The payout from a \$1 win bet calculated as $((WinAmount/Winners) \times Won) - 1$ . We also calculate realized returns analogously for place (top-two finish) and show (top-three finish) bets in the subsample for which data on these bets' odds are available.   |
| Position                        | The post-race finishing position of a horse where 1 indicates first place.   |
| Won                             | The post-race outcome that takes the value of 1 if a horse wins a race and 0 otherwise.  |
| <b>Odds</b>                     |  |
| Decimal Odds                    | The starting decimal odds for a specific horse.  |
| Win amount                      | The payout from a \$1 win bet; calculated as $1/DecimalOdds$ .   |
| Fractional odds                 | The starting odds in fractional form e.g., median fractional odds of "9-1" in Table 1 represent median decimal odds of 0.10); calculated as $WinAmount - 1$ .  |
| Fractional odds (log)           | The natural log of fractional odds.  |
| Odds-predicted position         | The finishing positions predicted by sorting on starting odds.   |
| Nonfavorite                     | A binary indicator that takes the value of 1 if fractional odds are greater than the sample median, and 0 otherwise. We consider alternate definitions in Table 5.   |
| <b>Trainer surname category</b> |  |
| Non-White                       | A binary indicator that takes the value of 1 if a trainer's surname is greater than 20% Hispanic, greater than 20% Black, greater than 20% Black or Hispanic, or missing using percentages from the U.S. Census Bureau's "Frequently Occurring Surnames from the 2010 Census" file; and 0 otherwise. We consider alternate definitions in Table 6. |
| Hispanic                        | A binary indicator that takes the value of 1 if a trainer's surname is greater than 20% Hispanic and less than 20% Black; and 0 otherwise.   |
| Black                           | A binary indicator that takes the value of 1 if a trainer's surname is greater than 20% Black and less than 20% Hispanic; and 0 otherwise.   |
| Ambiguous                       | A binary indicator that takes the value of 1 if a trainer's surname is greater than 20% Hispanic and 20% Black, or missing from the "Frequently Occurring Surnames from the 2010 Census" file; and 0 otherwise.  |
| <b>Horserace attributes</b>     |  |
| Runners                         | The total number of horses running in a race.  |
| Purse                           | The dollar amount that the race is worth.  |
| Horserace                       | A unique identifier for each specific race.  |

**TABLE A.2** | Realized returns by odds and trainer surname race/ethnicity.

|                         | Dependent variable: Realized return (\$) |                        |                        |                        |                        |                        |                        |                        |
|-------------------------|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                         | (1)                                      | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Constant                | -0.2586***<br>(0.0145)                   | -0.1522***<br>(0.0235) | -0.1417***<br>(0.0269) | -0.1137***<br>(0.0249) | -0.0978***<br>(0.0274) | -0.1528***<br>(0.0274) | -0.1563***<br>(0.0260) | -0.2216***<br>(0.0323) |
| Hispanic                | 0.1325**<br>(0.0594)                     | 0.1534***<br>(0.0591)  | 0.1539***<br>(0.0591)  | 0.0149<br>(0.0422)     | -0.0377<br>(0.0631)    | -0.0163<br>(0.0487)    | —                      | —                      |
| Black                   | 0.1176***<br>(0.0399)                    | 0.1186***<br>(0.0399)  | 0.1190***<br>(0.0399)  | 0.0841***<br>(0.0288)  | 0.1103***<br>(0.0396)  | 0.0865***<br>(0.0310)  | —                      | —                      |
| Ambiguous               | 0.1002**<br>(0.0420)                     | 0.0962**<br>(0.0421)   | 0.0964**<br>(0.0421)   | 0.0064<br>(0.0294)     | -0.0232<br>(0.0432)    | 0.0364<br>(0.0324)     | —                      | —                      |
| Odds (log)              |  | -0.0492***<br>(0.0126) | -0.0645***<br>(0.0194) | -0.0670***<br>(0.0192) | -0.0766***<br>(0.0196) | -0.0476**<br>(0.0199)  | -0.0333*<br>(0.0200)   | -0.0062<br>(0.0220)    |
| Nonfavorite             |  |                        | 0.0461<br>(0.0376)     | 0.0004<br>(0.0421)     | 0.0108<br>(0.0463)     | -0.0100<br>(0.0429)    | -0.0266<br>(0.0426)    | -0.0218<br>(0.0481)    |
| Hispanic × Nonfavorite  |  |                        |                        | 0.2353**<br>(0.1024)   | 0.3030***<br>(0.1112)  | 0.2518**<br>(0.1046)   | 0.2865**<br>(0.1159)   | 0.3837***<br>(0.1294)  |
| Black × Nonfavorite     |  |                        |                        | 0.0713<br>(0.0808)     | 0.0483<br>(0.0890)     | 0.0874<br>(0.0824)     | 0.0520<br>(0.0861)     | 0.0702<br>(0.0970)     |
| Ambiguous × Nonfavorite |  |                        |                        | 0.1917**<br>(0.0887)   | 0.2292**<br>(0.0957)   | 0.1871**<br>(0.0897)   | 0.1847**<br>(0.0923)   | 0.1887*<br>(0.1007)    |
| Horserace FE            |  |                        |                        |                        | ✓                      |                        |                        | ✓                      |
| Jockey FE               |  |                        |                        |                        |                        | ✓                      |                        | ✓                      |
| Trainer FE              |  |                        |                        |                        |                        |                        | ✓                      | ✓                      |
| N                       | 74,988                                   | 74,988                 | 74,988                 | 74,988                 | 74,988                 | 74,689                 | 74,156                 | 73,856                 |

Notes: The table presents results from estimation of variations of Equation (3). The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

**TABLE A.3** | Outcomes by morning line odds and trainer surname category.

|                        | (1)                    | (2)                    | (3)                    | (4)                   | (5)                    | (6)                    |
|------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|------------------------|
|                        | <b>Realized return</b> |                        |                        |                       |                        |                        |
|                        | <b>Actual odds</b>     | <b>M/L odds</b>        | <b>Position</b>        |                       | <b>Won</b>             |                        |
| Constant               | -0.1965***<br>(0.0396) | 0.0012<br>(0.0358)     | 2.2705***<br>(0.0300)  | 2.4299***<br>(0.0308) | -0.0081***<br>(0.0026) | -0.0333***<br>(0.0033) |
| Non-White              | 0.1667***<br>(0.0482)  | 0.1033***<br>(0.0335)  | -0.0768***<br>(0.0296) | -0.0651**<br>(0.0305) | 0.0077**<br>(0.0037)   | 0.0087**<br>(0.0038)   |
| Odds (log)             | -0.0295<br>(0.0210)    |                        |                        |                       |                        |                        |
| M/L Odds (log)         |                        | -0.1457***<br>(0.0192) |                        |                       |                        |                        |
| Predicted position     |                        |                        | 0.5428***<br>(0.0054)  |                       |                        |                        |
| M/L predicted position |                        |                        |                        | 0.5099***<br>(0.0056) |                        |                        |
| Odds (decimal)         |                        |                        |                        |                       | 0.8589***<br>(0.0170)  |                        |
| M/L odds (decimal)     |                        |                        |                        |                       |                        | 1.0364***<br>(0.0229)  |
| <i>N</i>               | 28,390                 | 28,390                 | 28,205                 | 28,205                | 28,390                 | 28,390                 |

*Notes:* The table presents results from estimation of Equation (2). The odd-numbered columns replicate the analysis from the analogous columns in Table 3 using the subsample for which morning line (M/L) odds data are available. The dependent variable in Column (1) is the realized return from a \$1 win bet. The dependent variable in Column (2) is the counterfactual realized return that would have been earned on a \$1 win bet if the starting parimutuel odds had been equal to the M/L odds. The dependent variables in Columns (3)–(6) are as in Table 3: finishing position of a horse (where 1 is first place and did-not-finish entries are dropped from the sample), and an indicator variable equal to 1 if the horse won the horserace. In all even-numbered columns, control variables based on odds are replaced with analogs calculated using M/L odds. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from the Racing Post–Horse Racing Nation matched sample and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

**TABLE A.4** | Realized returns by odds and White/non-White jockey surname.

|                                   | Dependent variable: Realized return (\$) |                        |                        |                        |                        |                        |                        |                        |
|-----------------------------------|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                                   | (1)                                      | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Constant                          | -0.2249***<br>(0.0304)                   | -0.1182***<br>(0.0356) | -0.1084***<br>(0.0381) | -0.1255***<br>(0.0331) | -0.1363***<br>(0.0452) | -0.1381***<br>(0.0255) | -0.1673***<br>(0.0392) | -0.2186***<br>(0.0323) |
| Jockey non-White                  | 0.0070<br>(0.0335)                       | 0.0020<br>(0.0336)     | 0.0022<br>(0.0336)     | 0.0225<br>(0.0264)     | 0.0501<br>(0.0407)     | —                      | 0.0160<br>(0.0327)     | —                      |
| Odds (log)                        |  | -0.0468***<br>(0.0126) | -0.0613***<br>(0.0194) | -0.0611***<br>(0.0194) | -0.0701***<br>(0.0199) | -0.0451**<br>(0.0200)  | -0.0323<br>(0.0200)    | -0.0049<br>(0.0220)    |
| Non-favorite                      |  |                        | 0.0438<br>(0.0376)     | 0.0754<br>(0.0658)     | 0.1101<br>(0.0722)     | 0.0895<br>(0.0682)     | 0.0535<br>(0.0689)     | 0.1037<br>(0.0784)     |
| Jockey non-White ×<br>nonfavorite |  |                        |                        | -0.0389<br>(0.0650)    | -0.0622<br>(0.0730)    | -0.0636<br>(0.0674)    | -0.0408<br>(0.0679)    | -0.0858<br>(0.0789)    |
| Horserace FE                      |  |                        |                        |                        | ✓                      |                        |                        | ✓                      |
| Jockey FE                         |  |                        |                        |                        |                        | ✓                      |                        | ✓                      |
| Trainer FE                        |  |                        |                        |                        |                        |                        | ✓                      | ✓                      |
| <i>N</i>                          | 74,988                                   | 74,988                 | 74,988                 | 74,988                 | 74,988                 | 74,689                 | 74,156                 | 73,856                 |

Notes: The table presents results from estimation of variations of Equation (3) using jockey surnames. There are 1272 unique jockey names in the sample. Three hundred and fifty-three jockeys have White surnames, while 919 have non-White surnames. Out of the non-White surnames, 676 are Hispanic, 129 are Black, and 114 are ambiguous (including 110 jockeys who appear in the horse racing logs but not the Census file). The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

**TABLE A.5** | Controlling for perceived trainer quality (across- and within-trainer).

|   | Dependent variable: Realized return (\$) |                        |                        |                        |
|---|--|------------------------|------------------------|------------------------|
|   | (1)                                      | (2)                    | (3)                    | (4)                    |
| Constant                                | -0.1127***<br>(0.0265)                   | -0.1227***<br>(0.0265) | -1.9188***<br>(0.0499) | -1.7313***<br>(0.0467) |
| Non-White                               | 0.0420**<br>(0.0210)                     | 0.0423**<br>(0.0208)   | -0.0856***<br>(0.0188) | —                      |
| Odds (log)                              | -0.0586***<br>(0.0199)                   | -0.0651***<br>(0.0192) | 0.3825***<br>(0.0239)  | 0.3438***<br>(0.0237)  |
| Nonfavorite                             | -0.0220<br>(0.0430)                      | -0.0011<br>(0.0420)    | -0.1246***<br>(0.0376) | -0.1435***<br>(0.0382) |
| Non-White × Nonfavorite                 | 0.1715***<br>(0.0588)                    | 0.1543***<br>(0.0565)  | 0.2269***<br>(0.0512)  | 0.1679***<br>(0.0553)  |
| Trainer won last race                   | -0.0771**<br>(0.0345)                    |                        |                        | -1.0372***<br>(0.0383) |
| Number of races in a year for a trainer |  | 0.0002<br>(0.0003)     |                        | -0.0004<br>(0.0004)    |
| Trainer win percentage for a track-year |  |                        | 0.0716***<br>(0.0014)  | 0.0735***<br>(0.0014)  |
| Trainer FE                              |  |                        |                        | ✓                      |
| N                                       | 72,073                                   | 74,988                 | 74,988                 | 71,675                 |

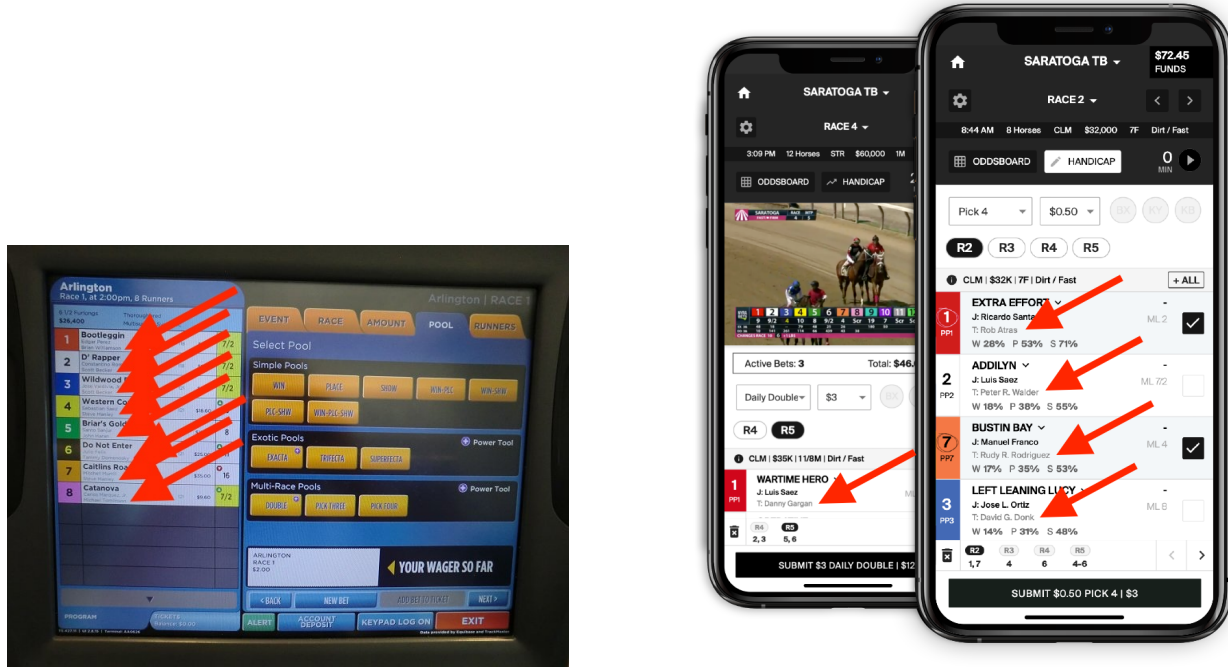
Notes: The table presents results from estimation of variations of Equation (3) using observable trainer controls and trainer fixed effects. The dependent variable is the realized return from a \$1 win bet. Coefficient estimates are shown with robust standard errors displayed in parentheses. The data are from 2011 to 2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

\*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively.

# Internet Appendix

Figure IA.1: Sample Sources of Information Available to Bettors

The figure illustrates various sources of information available to bettors, with the prominently featured trainer names indicated with red arrows.



(a) Terminal Bet

(b) Phone Bet

APPROX. POST: 4:24PM Daily Double-Exacta-Trifecta(5.50 min)-Pick 3(\$5.00 min) (Races 8-10-11)-Pick 4(\$5.00 min)(Races 9-12)-Superfecta(\$5.00 min)

**9** **KEENELAND**  
The 13th Running of The Madison (Grade I)

Win Place Show No.

Stakes: For fillies and mares, four years old and upward. By subscription of \$150 each, which should accompany the nomination or the Supplementary Nomination of \$150,000 at time of entry (includes entry and starting fees), \$7,500 to enter and an additional \$1,500 start, with \$300,000 guaranteed, of which \$150,000 to the owner of the winner, \$50,000 to second, \$30,000 to third, \$15,000 to fourth. \$50,000 to fifth and \$50,000 to be divided equally among sixth through last. 123 No winners of a Grade I stakes since November 1 allowed 3 lbs., a Grade II or III since October 4 or \$50,000 in 2014, 5 lbs. The maximum number of starters for the Madison will be limited to fourteen. In the event that more than fourteen fillies or mares appear for the entry box, the business stakes will be determined at that time with performance given to graded stakes winners in order (1st-10th), then highest 2013-2014 earnings. Starters to be named through the entry box by the usual time of closing. Keeneland will present a trophy cup to the winning owner. A silver jockey cap will be presented to the winning trainer and jockey. Closed Wednesday, March 26, 2014 with 34 nominations. Seven Furlongs

Track Record: Informed Decision (3), 1:17.86 (15-18-56)

| Program  | Class | Owner          | Trainer        | Jockey         | Class  |
|--|-------|----------------|----------------|----------------|--|
| 1  | 8-5   | Wesley A. Ward | John Velazquez | John Velazquez | 147-85 2014 1 1 0 0 \$60,000 Turf 2 1 1 8 \$38,775 |
| <p><b>Red</b> <b>Judy the Beauty (L)</b> 120<br/>                     Sire: Black Gold WW bet good on omelets, good WW hoop on black cap<br/>                     Trainer: Charles (1-2-1) 100.00%</p>   |       |                |                |                |  |
| <p><b>2</b> 4-1<br/>                     Sire: Orange, orange cap<br/>                     Trainer: David N. Brannaman and Paul Frazier</p> <p><b>White</b> <b>Apropos (L)</b> 120<br/>                     Sire: B's First Samurai - Apt by A.P. Indy - Breed in Kentucky by Claiborne Farm &amp; A&amp;B's - Discontinued (Apr 21, 2009)<br/>                     Trainer: Peter Miller (1-0-0-0) 100.00%</p>        |       |                |                |                |  |
| <p><b>3</b> 15-1<br/>                     Sire: Lurgone, black and red coat<br/>                     Trainer: Peter Miller (1-0-0-0) 100.00%</p> <p><b>Blue</b> <b>Heir Kitty (L)</b> 123<br/>                     Sire: B's First Samurai - Apt by A.P. Indy - Breed in Kentucky by Claiborne Farm &amp; A&amp;B's - Discontinued (Apr 21, 2009)<br/>                     Trainer: Peter Miller (1-0-0-0) 100.00%</p> |       |                |                |                |  |

9-12-2014 Race 9

RACE 9 CONTINUED ON NEXT PAGE

(c) Paper Race Card

Figure IA.2: Realized Returns by Trainer Surname Category

The figure plots average realized returns from \$1 win bets by surname categorization (dots) with 95% confidence intervals (vertical lines) from regressions of realized returns on indicator variables for each indicated trainer surname category. The dashed horizontal line at approximately  $-\$0.22$  shows the average realized return in the sample. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

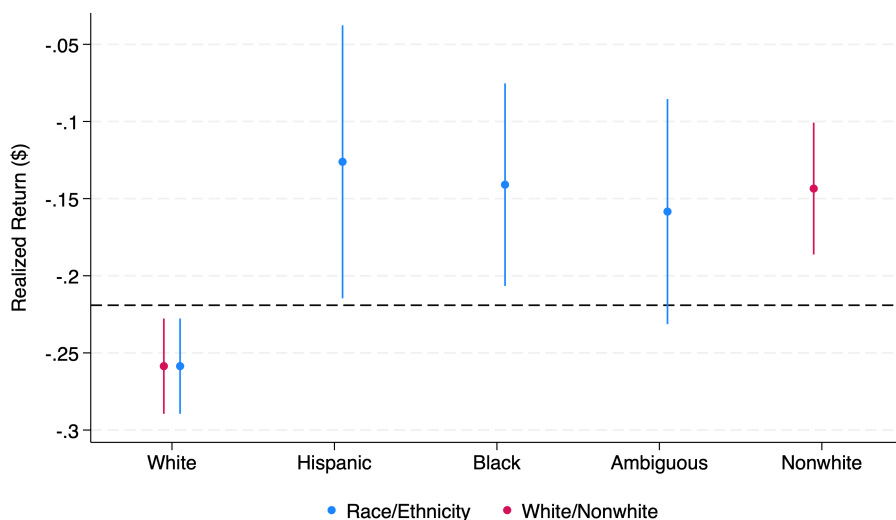


Figure IA.3: Odds Distribution by White/Nonwhite Trainer Surname

The figure graphs distributions of fractional odds by white/nonwhite categorizations. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

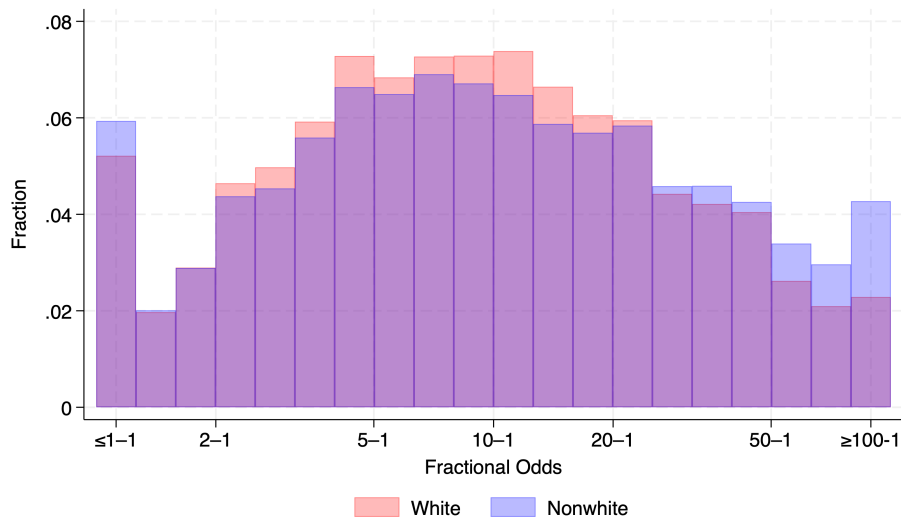


Figure IA.4: Realized Returns by Odds and Trainer Surname Category

The figure graphs local polynomial fits (and 95% confidence intervals) for realized returns from \$1 win bets and fractional odds on a log odds scale with a bandwidth of one log odds (as in Figures 3 and 4), separately by trainer surname race/ethnicity. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

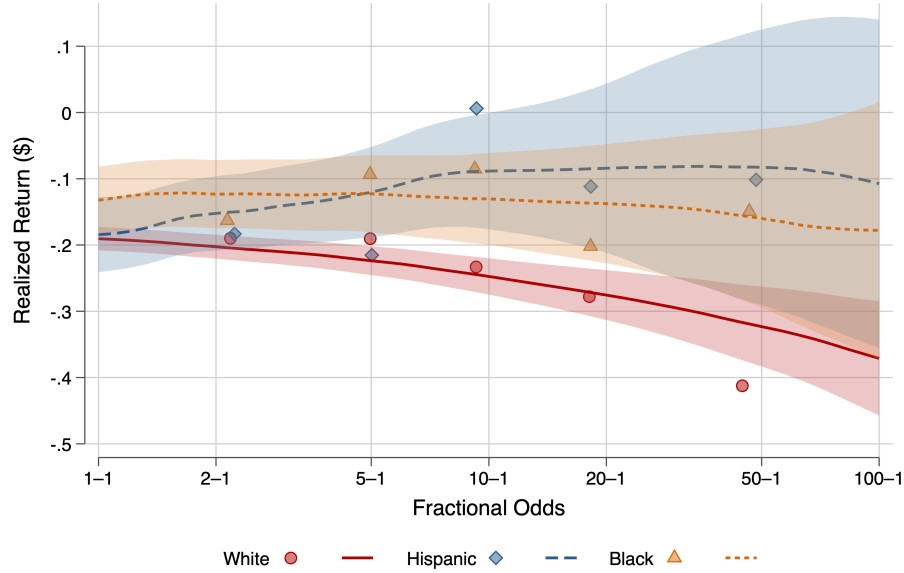


Table IA.1: Horse Trainers with Highest Race/Ethnicity Surname Percentages

The table lists the ten trainers with the highest race/ethnicity surname percentages. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1.

| Surname % White    |       | Surname % Hispanic |      | Surname % Black  |      |
|--------------------|-------|--------------------|------|------------------|------|
| Chris Stenslie     | 100.0 | Giovanni Luqueno   | 99.0 | Ricky Demouchet  | 96.1 |
| Anthony Granitz    | 100.0 | Cirilo Gorostieta  | 98.5 | Derrick Parram   | 90.4 |
| Bruce Kravets      | 99.7  | Olivo Inirio       | 97.8 | Paul Darjean     | 88.4 |
| Andrew Bossung     | 99.4  | Fernando Bahena    | 97.7 | Lawrence Bushrod | 85.0 |
| Leah Gyarmati      | 99.1  | Sergio Ledezma     | 97.1 | Larry Demeritte  | 77.0 |
| Paul Holthus       | 99.0  | Laura Cazares      | 96.6 | Hubert Pinnock   | 74.8 |
| Lee Couchenour     | 99.0  | Johanna Urieta     | 96.4 | Barbara Heads    | 74.1 |
| Danny Pish         | 98.9  | Henry Argueta      | 96.2 | John Ivory       | 73.2 |
| Vincent Moscarelli | 98.8  | Alfredo Velazquez  | 96.0 | Joseph Cheeks    | 73.1 |
| Cathy Denelsbeck   | 98.8  | Cesar Nambo        | 96.0 | Aubrey Maragh    | 71.6 |

## Heterogeneous effects in the U.S. South

Local histories and demographics may lead to differential patterns of racial/ethnic pricing gaps (as, for example, in Doleac and Stein, 2013 and Dougal et al., 2019). In Table IA.2, we replicate our main analysis separately for horseraces in and outside the U.S. South (which represents nearly half of our sample). The resulting estimates follow a somewhat similar pattern as the heterogeneous effects by track condition. While white–nonwhite return gaps are similar in and outside the South on average (as shown in columns 3 and 5), Southern gaps are entirely driven by longshots (as shown in column 4).<sup>26</sup> Outside the South, the differential effects of race/ethnicity on returns vary less between favorites and longshots. Although data does not of course allow us to measure any particular bettors’ biases, our results are consistent with the possibility that discriminatory attitudes are more heterogeneous in the South, with higher levels of bias particularly concentrated among less-sophisticated bettors there.

Table IA.2: Realized Returns in the U.S. South

The table presents results from estimation of variations of equation 3. The dependent variable is the realized return from a \$1 win bet. Columns (1) and (2) report pooled regressions using the full sample of observations. Columns (3) and (4) show the sub-sample of horses that run in horseraces at a track located in the Southern U.S. Columns (5) and (6) contain the sub-sample of horses that run in horseraces at a track located outside the Southern U.S. Coefficient estimates are shown with robust standard errors displayed in parentheses. \*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

|                         | Dependent Variable: Realized Return (\$) |                        |                        |                        |                        |                        |
|-------------------------|--|------------------------|------------------------|------------------------|------------------------|------------------------|
|                         | (1)                                      | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    |
|                         | Pooled                                   |                        | South                  |                        | Non-South              |                        |
| Constant                | -0.1438***<br>(0.0269)                   | -0.1146***<br>(0.0248) | -0.1432***<br>(0.0398) | -0.0996***<br>(0.0382) | -0.1447***<br>(0.0366) | -0.1295***<br>(0.0320) |
| Nonwhite                | 0.1193***<br>(0.0285)                    | 0.0423**<br>(0.0208)   | 0.1220***<br>(0.0424)  | -0.0022<br>(0.0311)    | 0.1173***<br>(0.0384)  | 0.0795***<br>(0.0280)  |
| Odds (log)              | -0.0634***<br>(0.0193)                   | -0.0663***<br>(0.0192) | -0.0580**<br>(0.0293)  | -0.0611**<br>(0.0292)  | -0.0681***<br>(0.0254) | -0.0699***<br>(0.0250) |
| Non-Favorite            | 0.0459<br>(0.0376)                       | -0.0010<br>(0.0420)    | 0.0172<br>(0.0561)     | -0.0546<br>(0.0610)    | 0.0725<br>(0.0503)     | 0.0480<br>(0.0582)     |
| Nonwhite × Non-Favorite |  | 0.1532***<br>(0.0565)  |                        | 0.2285***<br>(0.0799)  |                        | 0.0818<br>(0.0801)     |
| <i>N</i>                | 74,988                                   | 74,988                 | 36,970                 | 36,970                 | 38,018                 | 38,018                 |

<sup>26</sup>In Table IA.3, we find that discrimination in the South is mostly associated with return differences for Hispanic-named trainers. Outside of the South, discrimination appears to be strongest against Black-named trainers.

Table IA.3: Realized Returns by Odds and Trainer Surname Race/Ethnicity in the U.S. South

The table presents results from estimation of variations of equation 3. The dependent variable is the realized return from a \$1 win bet. Columns (1) and (2) report pooled regressions using the full sample of observations. Columns (3) and (4) show the sub-sample of horses that run in horseraces at a track located in the Southern U.S. Columns (5) and (6) contain the sub-sample of horses that run in horseraces at a track located outside the Southern U.S. Coefficient estimates are shown with robust standard errors displayed in parentheses. \*\*\*(\*\*)(\*) indicate significance at the 1%(5%)(10%) two-tailed levels, respectively. The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Table A.1.

|                                 | Dependent Variable: Realized Return (\$) |                        |                        |                       |                        |                        |
|---------------------------------|--|------------------------|------------------------|-----------------------|------------------------|------------------------|
|                                 | (1)                                      | (2)                    | (3)                    | (4)                   | (5)                    | (6)                    |
|                                 | Pooled                                   |                        | South                  |                       | Non-South              |                        |
| Constant                        | -0.1417***<br>(0.0269)                   | -0.1137***<br>(0.0249) | -0.1358***<br>(0.0402) | -0.0949**<br>(0.0385) | -0.1496***<br>(0.0362) | -0.1312***<br>(0.0319) |
| Hispanic                        | 0.1539***<br>(0.0591)                    | 0.0149<br>(0.0422)     | 0.2639***<br>(0.0799)  | -0.0097<br>(0.0567)   | -0.0088<br>(0.0879)    | 0.0384<br>(0.0632)     |
| Black                           | 0.1190***<br>(0.0399)                    | 0.0841***<br>(0.0288)  | 0.0583<br>(0.0567)     | 0.0145<br>(0.0412)    | 0.1901***<br>(0.0560)  | 0.1469***<br>(0.0404)  |
| Ambiguous                       | 0.0964**<br>(0.0421)                     | 0.0064<br>(0.0294)     | 0.0823<br>(0.0712)     | -0.0236<br>(0.0504)   | 0.1054**<br>(0.0523)   | 0.0298<br>(0.0366)     |
| Odds (log)                      | -0.0645***<br>(0.0194)                   | -0.0670***<br>(0.0192) | -0.0620**<br>(0.0295)  | -0.0648**<br>(0.0295) | -0.0664***<br>(0.0252) | -0.0684***<br>(0.0249) |
| Non-Favorite                    | 0.0461<br>(0.0376)                       | 0.0004<br>(0.0421)     | 0.0202<br>(0.0564)     | -0.0472<br>(0.0615)   | 0.0749<br>(0.0504)     | 0.0452<br>(0.0580)     |
| Hispanic $\times$ Non-Favorite  |  | 0.2353**<br>(0.1024)   |                        | 0.4507***<br>(0.1368) |                        | -0.0691<br>(0.1541)    |
| Black $\times$ Non-Favorite     |  | 0.0713<br>(0.0808)     |                        | 0.0852<br>(0.1074)    |                        | 0.0980<br>(0.1243)     |
| Ambiguous $\times$ Non-Favorite |  | 0.1917**<br>(0.0887)   |                        | 0.2042<br>(0.1379)    |                        | 0.1748<br>(0.1161)     |
| <i>N</i>                        | 74,988                                   | 74,988                 | 36,970                 | 36,970                | 38,018                 | 38,018                 |