Racial discrimination in asset prices: Evidence from horse betting*

Spencer Barnes
University of Texas at El Paso

Luke C.D. Stein
Babson College

January 17, 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 non-financial 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 non-pecuniary returns (due to animus). Bets on horses with nonwhite-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, in low-stakes races with "fast" conditions, and in the U.S. South. These results are consistent with the effects of discrimination being strongest among less-informed and less-sophisticated bettors.

JEL Codes: G14, G41, J15, J71, L83, Z20, Z23

*Barnes: Dept. of Economics and Finance, Woody L. Hunt College of Business Administration, The University of Texas at El Paso; scbarnes@utep.edu. Stein: Finance Division, Babson College; lcdstein@babson.edu.
We would like to thank Alex Binder, Bruce Carlin, Will Cassidy, Tony Cookson, Ryan Davies, Ashleigh Eldemire-Poindexter, Etan Green, Paul Grimes, Scott Guernsey, Jarrad Harford, Brandon Mendez, Justin Mohr, Andy Puckett, Tian Qiu, Laura Schechter, Clemens Sialm, Kaushik Vasudevan, and James Weston for useful feedback, along with participants in the 2023 Lone Star Finance Conference and 2024 Univ. of Washington Fostering Inclusion DEI Research Workshop. This paper supersedes an earlier working paper circulated with the title "Racial Biases in Prices: Fair Pricing Distortions at the Horse Track."

"It's a lot of racism. A lot.... They make it seem like something's wrong with the horse and there's nothing you can do about it."

—Uriah St. Lewis, Horse trainer

"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 non-financial. 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 nonwhite-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 U.S. with around $12 billion bet in 2022, or about 14% of the total amount wagered by U.S. sports bettors (Fenly, 2023; Yakowicz, 2023). Horses race throughout the U.S., with a typical Saturday during peak racing season (May–June) seeing around 30 tracks each holding 8–12 races with an average of 9 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.[1] 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.[2]

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 A.1.

[1]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).

[2]When betting is *not* parimutuel, winning bettors typically receive different payouts that depend on what platform they use to place their bets and when those 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 where officials exercise limited discretion. 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).

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 the jockey 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.[3] This could include race or ethnicity, which bettors presumably infer—consciously or not—from trainers' surnames, which are visible and highly salient.[4]

Some bettors may also have non-pecuniary reasons for preferring horses with trainers who are members of 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 also to support her pre-existing belief in white trainers' superiority. Although we cannot directly observe bettors' assessments of trainer quality or their 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.[5]

[3]Limited attention likely also plays a role, given the large number of horse trainers active in the U.S., and the large number of horses racing on a given day. Bettors may not have the inclination or ability to assess quality directly, and therefore rely on heuristics to make decisions.

[4]While surnames are highly salient, bettors are not typically exposed to other signals of trainer demographics (in person, through media appearances, or in handicapping outlets) 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 might praise as handicapping can be derided as mere superstition by others, echoing attempts to differentiate "accurate" from "inaccurate statistical discrimination" in other economic contexts.

[5]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

We analyze bets on 74,988 entrants in 9,164 U.S. horseraces held between 2011 and 2022, focusing on the relationship between parimutuel odds, realized returns, and trainers' surnames.[6] We assess the extent to which a trainer might be perceived by bettors as likely to be nonwhite using the frequency of surnames in the 2010 U.S. Decennial Census. Horses with nonwhite-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 nonwhite-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 nonwhite-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 non-pecuniary, recreational tolerance for risk-taking (Feess et al., 2014; Geertsema and Schumacher, 2016). These gamblers should therefore play a particularly important role in

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.

[6] Our main sample is described in detail in Section 3.1, and uses data from Racing Post, and largely covers "stakes races."

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 nonwhite-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 nonwhite-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 the "place" and "show" bets (where bettors succeed if their horse finishes in the top two or three, respectively) 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 those 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.[7] Binder et al. (2021) and Binder and Grimes (2022) find evidence that female jockeys in the U.S. underperform relative to bettor expectations, consistent with discriminating in favor of females, while evidence from the U.K. suggests female jockeys exhibit relative overperformance there (Brown and Yang, 2015; Cashmore et al., 2022). Leeds

[7]We control for any differences in the jockeys hired by white and nonwhite-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 Appendix Table A.5. Unlike trainers—who play a long-term strategic role in determining horse performance—U.S. jockeys are overwhelmingly non-white (approximately 72% using our methodology in our sample) and we do not find evidence for racial/ethnic discrimination here. While anecdotal evidence suggests that superstitious bettors may pay attention to horse names, these are not typically racialized.

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.[8]

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 integrate the horse racing discrimination and FLB literatures, explicitly considering asset returns and introducing a set of tests that consider 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 nonwhite-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. 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), and art and collectibles (Adams et al., 2021; Kim and Lee, 2022). 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 approach suggest that discrimination may have the largest effects where limits to arbitrage are strong and investors

[8]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 U.K. where implied win probabilities are calculated using prices from various bookmakers.

are unsophisticated or poorly informed, hypotheses worth investigating in other markets.

2 Horse Racing Primer

Horse racing is a storied pastime; records indicate that ancient Greeks and Romans would spend a day at the races. Europe has typically been the heart of horse racing in more modern eras, with one of the most prominent racetracks—Newmarket, in Suffolk, England—founded during the rule of King James I in 1636.

The earlier records of mass betting on horse racing occur during the 1600s in England. European kings were such avid sports fans that horse racing flourished until the 1900s, when attention shifted to the U.S. Even though the sport lost some interest in the 2000s, the betting markets remain as active as ever.

In the 2024 running of the Kentucky Derby alone, there was a total of $211 million in bets, a new record for the race (Purdum, 2024). The U.S. horse racing market is estimated to run about $12 billion annually in legal betting (Fenly, 2023). Estimates suggest 2023 betting volumes of around $18 billion in Hong Kong (HKJC, 2023), and $1 billion in the U.K. (U.K. Gambling Commission, 2024).

Among betting markets, horse races provide one of the closest analogs to more traditional financial asset markets. In the U.S., the specific structure is parimutuel betting.[9] 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.

Although odds are updated as bets are placed, payoffs do not depend on the odds posted *when* a bet is placed. Payoffs are determined only by the total amount bet as of the time the race starts, called the starting odds. Starting odds are determined by the demand for bets on each horse in a race and, therefore, influenced by any factors that affect this demand.

[9]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 on whether a horse will finish in the top two or three positions, as well as more exotic bets.

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 betters 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.

[Insert Figure 1 Here]

Figure 1 shows an example of a race card entry. This is standard information that is readily available to all bettors. The card is for race seven at Oaklawn Park in Hot Springs, Arkansas, set to go off at 3:54 p.m. The race allows win, place, and show bets, along with a variety of "exotic" bets (daily double, exacta, trifecta, and superfecta). 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 Appendix Figure A.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

This paper's two primary data sources are drawn from U.S. horse racing logs and the U.S. Census Bureau.

Our horse racing data comes from web-scraping publicly available horse racing websites and covers the period from 2011 to 2022. The data is from open public websites without sign-ups, logins, or other restrictions. Specifically, our main analysis sample is built using information scraped from `racingpost.com`, which overwhelmingly covers stakes races in the U.S. This sample includes 9,164 horseraces and 74,988 total entrants.

In some analysis, we supplement Racing Post data with information on morning line odds and place/show betting pools. This data is scraped from `horseracingnation.com`, which is also available on the public web. Horse Racing Nation data are user-generated and only cover the period back to 2018. We match horses and horseraces from Racing Post to Horse Racing Nation for the years where they overlap (2018–2022).

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 cleanly identify how bias affects pricing outcomes. Surnames are ingrained within families and 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).

[Insert Figure 2 Here]

The U.S. Census Bureau's "Frequently Occurring Surnames from the 2010 Census" zipped data package contains a demographic breakdown of the full-count population of individuals. It contains 162,253 surnames from the 2010 U.S. Decennial Census. The U.S. horse racing logs from Racing Post contain 2,915 unique trainers, of whom 2,634 have surnames that appear in the Census file. Figure 2 plots these matched trainer surnames by the fraction of the overall population having that surname who are classified as Hispanic and Black.[10] For each trainer (and therefore each race entrant), we create categorical variables reflecting whether their name is likely perceived as white, Hispanic, Black, or ambiguous; the latter three categories we also label collectively as "nonwhite."

We impose cuts at both 20% Hispanic and Black Census surname percentages, indicated by horizontal and vertical lines. (To ensure that these thresholds are not not solely responsible for our findings, we also consider robustness to several alternative trainer surname categorizations.) Each quadrant corresponds to a race/ethnic categorical variable, with observations in the lower-left categorized as white, the upper-left as Black, the lower-right as Hispanic, and the upper-right as ambiguous. In addition, the 281 unplotted trainers whose uncommon surnames mean they do not appear in the Census file are also categorized as ambiguous.[11]

3.2 Summary Statistics

Table 1 shows summary statistics for several key horserace and bet attributes and outcomes, as well as information about our baseline categorization of trainers' surnames. Nonwhite-named trainers account for 35% of horse-race level observations; 8%, 15%, and 12% are categorized as Hispanic, Black, and ambiguous, respectively. The median finishing position is fifth place (mean 4.9), necessarily in the middle of the pack among the median horserace's

[10] We follow the AP Stylebook in capitalizing "Black" and "Hispanic" but not "white."

[11] To better conceptualize the U.S. Census data's perceived race/ethnicity measure, Appendix Table A.2 lists the sample's most racially/ethnically distinctive trainer surnames. The first two columns show perceived white trainers; the second two columns list perceived Hispanic trainers; and the third two columns list perceived Black trainers. The match suggests that "Denelsbeck" is a predominately white surname, with 99% of the U.S. population having that surname classified as white in the U.S. Census. Analogously, "Velazquez" is an overwhelmingly Hispanic surname (96% classified as Hispanic) and Ivory is a highly Black surname (73% classified as Black). White and Hispanic trainers' perceived race/ethnicity is much more certain than that of Black trainers, as the percentages are closer to 100%. In contrast, Black trainers have lower surname percentages, around 75%.

nine runners (mean 8.9). About 12% of horses win a horserace, so the median win bet doesn't pay out and earns a realized return of -100%.[12]

[Insert Table 1 Here]

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. Using the starting (decimal) odds from the U.S. horse racing logs, Table 1 shows that the average decimal odds in our sample are 0.15; horserace 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.[13]

The maximum potential (gross) payout for a \$1 win bet on a given horse can be calculated from their decimal odds as $WinAmount = 1/DecimalOdds$. Odds 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 post-race bet performance, assessing the realized (net) return as a function of *WinAmount* and the race results, using an indicator for whether the horse won the race (*Won*), and accounting for the rare occurrence of "dead heats" (effectively ties, characterized by $Winners > 1$) where the parimutuel pool is shared

[12] All summary statistics are reported over the sample of horses × horseraces, so trainers are effectively weighted by the total number of times their horses race, and horseraces are effectively weighted by their number of runners.

[13] 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.

across bets on multiple horses:

$$RealizedReturn = WinAmount \cdot \frac{Won}{Winners} - 1.$$

Table 1 shows that the average realized return in the sample is –\$0.22.[14] This corresponds to the –22% loss suffered by a bettor who held an equal-weighted portfolio by placing an equally-sized win bet on each horse in each race. The negative mean is driven by the track's "takeout" from each parimutuel pool (although the aggregate size of the takeout corresponds to the return on a *value-weighted* portfolio and is typically less than 22%). 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 horses with lower fractional odds (favorites) generate higher average realized returns than those with higher fractional odds (longshots) (Thaler and Ziemba, 1988). Alternate statements of the FLB are that longshots are "overbet," or that on average they underperform bettors' expectations. There have been many explanations offered for the FLB, of which Ottaviani and Sørensen (2008) provide a useful review.[15] While explaining the FLB is outside the scope of this paper, we nonetheless document the empirical regularity in our sample in Figure 3.

[14]This is close to the average realized returns of –\$0.23 in the Snowberg and Wolfers (2010) and –\$0.25 in the Green et al. (2020) samples. Track takeouts vary somewhat over time and across tracks; our slightly less negative average return may also relate to the Racing Post sample's heavy weight on stakes races.

[15]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, non-expected-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.

[Insert Figure 3 Here]

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 mean returns, 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, which is 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 takeaway from the FLB is that bettors could reduce their losses by wagering on favorites instead of longshots.

We model the FLB in regression form as

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

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 non-favorite based on offering above-median odds; the interaction $(\log(FO_{hr}) \times NF_{hr})$, which allows slopes to vary across the favorite/non-favorite part of the distribution; and ω_r, horserace fixed effects. ϵ_{hr} is the econometric error term, and we report robust standard errors.

[Insert Table 2 Here]

The results from OLS estimates of equation 1 are reported in Table 2. The columns of the table relax successive constraints that the regression coefficients be zero. Column (1) reports the average realized return of the sample, which is –0.22 and is statistically significant at the 1% level. Intuitively, a \$1 win bet will lose \$0.22. 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 twelve cents. For a one-unit increase in log fractional odds, there is a corresponding 4.7 percentage point decrease in realized returns.

However, the relationship in Figure 3 exhibits nonlinearity, with what appears to be a kink near the sample median (fractional odds around 9–1). A similar, stark break in the slope of the odds–return relationship is illustrated by Snowberg and Wolfers (2010, Fig. 1) in their sample. We therefore allow in column (3) a break in the level of the realized returns-to-odds relationship at this median breakpoint by including an indicator variable *Non-Favorite* 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.[16] 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 non-favorite part of the odds 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 non-favorites, will inform our approach for understanding the relationship between race/ethnicity and returns across the risk distribution. Many theoretical 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). Lower average returns on longshots should disproportionately attract less-informed bettors and those with non-pecuniary, recreational tolerance for risk-taking. Feess et al. (2014) and Geertsema and Schumacher (2016) show that inexperienced and unsophisticated indeed bet more on longshots. These gamblers

[16] We do not find statistically significant evidence for a level break at median odds, which corresponds to the *Non-Favorite* coefficient in column (3). In columns (4)–(5), this coefficient represents differences in the intercept, rather than the size of a level break (if there were one).

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.

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 Thoursie, 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 nonwhite-named trainers horses' finishing positions.[17] We therefore focus on assessing bettor-driven differences in financial outcomes and start to consider here how surnames might impact horse pricing outcomes (Barghouty et al., 2020; Feigenberg and Miller, 2022).

[Insert Figure 4 Here]

Figure 4 plots the average realized returns for $1 win bets by race/ethnicity, and their 95% confidence intervals. Indeed, Figure 4 suggests that trainers' perceived race/ethnicity influences pricing decisions, as considerable heterogeneity exists among realized returns; the average realized return for white-named trainers is –$0.26, worse than the sample average realized return of –$0.22. On average, nonwhite-named trainers have better realized returns, with a return of –$0.15. The white–nonwhite difference is significant at the 1% level, as shown in Table 4, column (1).

[17] Mean finishing position for white vs. non-white: 4.90 vs. 4.91, t-test for unequal means $p = 0.54$. Standard deviation: 2.88 vs. 2.88, F-test for unequal variances $p = 0.75$. Wilcoxon rank-sum (Mann–Whitney) test: $p = 0.32$.

Hispanic-named trainers primarily drive this difference, with average realized returns of –\$0.13. These are less negative (though not statistically significantly) than Black- and ambiguously-named trainers' averages of approximately –\$0.14 and –\$0.16, respectively. Each of the three nonwhite-named groups offer average returns that exceed the white-named trainers' by amounts significant at the 5% and 1% levels (as shown in Appendix Table A.3, column 1).

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 nonwhite-named trainers could be generated mechanically by the favorite–longshot bias if these trainers disproportionately trained favorites. However, this is not the case.

[Insert Figure 5 Here]

Figure 5 plots the odds distribution separately for white- and nonwhite-named trainers. The data suggests that nonwhite-named trainers are more likely to train longshot horses, which receive lower average returns under the FLB (as shown in Figure 3 and Table 2).[18] The unconditional racial/ethnic return differences therefore go in the opposite direction of what the FLB alone would generate.

[Insert Table 3 Here]

In columns (1) and (2) of Table 3, we estimate returns by race/ethnicity as in Figure 4, but with a linear control for log fractional odds to account for the existence of the FLB and the different odds distributions across race/ethnicity. This gives regression models that are variations of

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

[18] Nonwhite-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.

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 nonwhite-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 consider 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 nonwhite-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 this effect exists across nonwhite 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 the race as the dependent variable. Here we replace log fractional odds with decimal odds as a control for win likelihood (with estimated coefficients, as expected, positive but less than one). Nonwhite-named trainers more horseraces than their decimal odds suggest, with Black-named trainers having the largest likelihood gaps from white-named trainers.

Taken together, the evidence presented in Table 3 shows that nonwhite-named (and particularly Hispanic-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 under-betting of these horses, either because bettors simply prefer not to bet on nonwhite-named trainers or because they (inaccurately) underestimate these trainers' quality.[19]

4.3 Realized Returns by Odds and Race/Ethnicity

Not only do *average* realized returns vary by trainer race/ethnicity as illustrated in Figure 4, but this variation is not constant between favorites and longshots. Figure 6 plots realized returns at quintiles of the odds distribution (with a local polynomial fit and its 95% confidence interval) separately for white- and nonwhite-named trainers, replicating Figure 3.

[Insert Figure 6 Here]

White-named trainers, who comprise approximately two-thirds of the sample, exhibit a pattern consistent with traditional estimates of the FLB as estimated in Section 4.1: Average realized returns for favorites have a nearly flat slope with respect to odds, with longer longshots offering increasingly negative returns starting near the median of the odds distribution. For example, 1–1, 5–1, and 10–1 favorites with white-named trainers all average

[19]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 or oddsmakers, 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 sub-sample for which M/L odds data is available. In Appendix Table A.4 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, in addition to realized returns on win bets, we analyze counterfactual returns that would have been earned if the starting odds were equal to those set by the track's morning line handicapper. We also consider racial/ethnic 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 also find (in untabulated regressions) that the differences between decimal M/L and final odds do not systematically vary by race/ethnicity. (This approach follows Igan et al., 2015, who assess the change in basketball point spreads as a function of teams' racial composition.) Our results suggests 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's suggestion that we consider this line of analysis.

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 nonwhite-named trainers. Consistent with the mean differences discussed in Section 4.2, horses with nonwhite-named trainers have higher average returns, and the sign of this difference is the same at every quintile of the odds distribution. (Figure 6 also shows that confidence intervals of separate local polynomial fits are also non-overlapping among all but the most extreme favorites and longshots.)

However, the white–nonwhite differences appear to be much larger among longshots than favorites. The pattern appears consistent with the existence of a bettor bias against nonwhite-named trainers' longshots, which is 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.[20]

We model these relationships 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 γ, on log fractional starting odds, and ξ, on an indicator for non-favorites (i.e., horses with above median odds), capture the FLB for white-named trainers. As in equation 2, the coefficient β on indicator variables for trainer name category captures realized return

[20]Appendix Figure A.2 plots realized returns as a smooth function of log fractional odds separately for white-, Hispanic-, and Black-named trainers. As in Figure 6, realized return averages for each odds quintile of Hispanic- and Black-named trainers are greater than white-named trainers (except for 5-1 Hispanic-named trainers). However, there is a more limited range of odds at which the nonparametric fits' confidence intervals are non-overlapping: Roughly 5–1 to 20–1 odds for Hispanic- and Black- versus white-named trainers. Also, similar to Figure 4, at no point in the range of odds are realized returns for Hispanic- and Black-named trainers statistically different from each other.

differences from white-named trainers, now specifically among favorites. The coefficient δ captures *additional* racial/ethnic return differences among non-favorites; these latter two are our key coefficients of interest. In some estimates, we include horserace fixed effects (ω_r), jockey fixed effects (ϕ_j), and/or trainer fixed effects (η_t) to control for other time-invariant characteristics.

[Insert Table 4 Here]

The results from estimating variants of equation 3 are shown in Table 4. Column (1) serves as a baseline model, including only the binary indicator variable for nonwhite-named trainer surnames as in Figure 4; the constant represents the average realized returns for white-named trainers across the entire odds distribution (a $0.26 loss on a $1 win bet). Nonwhite-named trainers offer an average return $0.12 higher. Column (2) adds the log fractional odds control (repeating column 1 of Table 3). In column (3) we allow a break in returns between favorites and non-favorites.

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 the risks in question are idiosyncratic, efficiency should even eliminate the return differences without risk adjustment.) We observe highly significant white–nonwhite differences of roughly 12 percentage points in columns (1) through (3).

In columns (4)–(8), we allow the white–nonwhite gap to differ between favorites and non-favorites, as suggested by Figure 6. 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–nonwhite return gap is 3–5 percentage points, and significant at the 5% level unless horserace fixed effects are included. In particular, the results are not driven by differences in the jockeys riding white- and nonwhite-trained horses.[21] The estimated interaction coefficients in these specifications show that the racial/

[21] We also replicate in Appendix Table A.5 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 nonwhite-named and white-named jockeys, perhaps because the latter are rare.

ethnic return gaps are 15–17 percentage points larger among non-favorites (differences significant at the 1% level).[22]

These results suggest that horse bettors discriminate against nonwhite-named trainers, and that this discrimination is concentrated among non-favorites. Given the existence of the favorite–longshot bias, non-favorites 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.[23]

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 nonwhite-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–nonwhite 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.[24] The estimated interaction coefficients show

[22]Similar results hold when we break nonwhite-named trainers out into Black and Hispanic in Appendix Table A.3. Discrimination is the strongest against Hispanic-named trainers among longshots. It is possible that discrimination is related to the racial/ethnic makeup of trainers participating in a given horserace. For example, bettors on a horserace where six of the seven trainers have white names might single out the sole nonwhite-named trainer's. The results presented in Appendix Table A.6 fail to find evidence that racial/ethnic return gaps vary systematically across within-race trainer surname diversity.

[23]We ensure the robustness of the results from Table 4 by varying our definitions of the measures of race/ethnicity and non-favorite. Appendix Table A.7 shows results when varying measures of race/ethnicity. In Panel (a), we move the Census surname cutoffs to 80% to create our binary categorization. In Panel (b), we use a standardized (mean zero, standard deviation one) continuous measure of nonwhite surname percentage. The results are similar to Table 4 in both panels. Appendix Table A.8 shows results also remain similar using alternate definition for non-favorites (within-horserace median, pooled 25th percentile, or quintiles).

[24]We also consider controls for several time-varying observable trainer characteristics in Appendix Table A.9. Column (1) includes Trainer Won Last Race, which is a binary variable that takes the value of one if a trainer has a horse that won their previous race and zero otherwise. This controls for the perceived "hot streak" of the trainer. Column (2) includes Number of Races in a Year for a Trainer, which is a continuous variable of the number of races a trainer has raced in that year up to the current race. We posit that this captures something like the "Baffert effect" where incredibly well-known and popular trainers have more horses in more races. Column (3) includes Trainer Win Percentage for a Track-Year, which is the number of races that a trainer has won at a specific track before the current race divided by the total number of horses that trainer has raced at that track up to that date in that year. This controls for something like a home-track effect. Lastly, column (4) includes all the observable controls together with trainer fixed effects. We find that

that the racial/ethnic return gap remains roughly 15–18 percentage points larger among non-favorites (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.

5 Heterogeneous Effects and Economic Mechanisms

Horse gambling is a market in which sophistication varies enormously across bettors (Roeder, 2023). We complement our results using variation in bettor 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.

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.

[Insert Table 5 Here]

The results for win, place, and show bets are in Table 5. We find that the our results for win bets hold in the matched Racing Post–Horse Racing Nation sample. Specifically, in column (1), win bets on nonwhite-named trainers earn 17 percentage point higher average

across the different columns, the estimated interaction coefficients show that the racial/ethnic gap is still between 15–23 percentage points larger among non-favorites, with differences significant at the 1% level.

returns than bets on white-named trainer. Column (4) shows that this effect is driven by longshots, where the white–nonwhite 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 who are more informed about a horse's recent performance 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.

[Insert Table 6 Here]

In Table 6 we present estimates of equation 3 (as in Table 4, columns 3–4), confirming our baseline results in the sub-sample for which we observe a previous finish, and partitioning the sample between horses that finished in the bottom versus the top half of their previous horserace. Figure 1 and Appendix Figure A.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 non-favorites. Comparing the estimates in columns (3) to (5) shows that among horses with a previous finish in the bottom half, nonwhite-named trainers generate 17 percentage point higher returns, almost triple the (statistically insignificant) 6 percentage point gap for top-half finishers. Column (4) shows that non-favorites 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 nonwhite-named trainers.

[Insert Table 7 Here]

Table 7 shows results split by purse size. Comparing columns (3) and (5) we find the magnitude of the average white–nonwhite return gap is slightly larger in small-purse horseraces (14 versus 10 percentage points). The results in columns (4) and (6) show that once again, the pricing effects of discrimination are driven by non-favorite horses, 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.

[Insert Table 8 Here]

In Table 8, we show estimates splitting between "fast" and other track conditions. While the average white–nonwhite return gap is an indistinguishable 11–12 percentage points on both types of track conditions (as shown in columns 3 and 5), we find that for fast conditions, this effect is entirely driven by bets on longshots. These horses and these conditions are precisely where we expect the least informed and least sophisticated bettors to have the strongest pricing effects, and are where we see evidence of the strongest racial/ethnic discrimination. Interestingly, with non-fast track conditions, we find statistically significant racial/ethnic pricing gaps among favorite horses, although the differences between favorites and longshots are smaller than for fast tracks (and statistically insignificant).

[Insert Table 9 Here]

Finally, we consider variation in effects across regions of the U.S. 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 9, 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).[25] Outside the South, the differential effects of race/ethnicity on returns vary less between favorites and longshots. Although data does not of course allow

[25]In Appendix Table A.10, 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.

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.

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 non-pecuniary returns. The results show that horses with nonwhite-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.

Risk-adjusting to control for the well-known "favorite–longshot" bias, the study finds that nonwhite-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, in low-stakes races with "fast" conditions, and in the U.S. South. 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.

References

Adams, R.B., Kräussl, R., Navone, M., Verwijmeren, P., 2021. Gendered prices. Review of Financial Studies 34, 3789–3839.

Aggarwal, R., Boyson, N.M., 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, 803–815.

Arai, M., Thoursie, P., 2009. Renouncing personal names: An empirical examination of surname change and earnings. Journal of Labor Economics 27, 127–147.

Asch, P., Malkiel, B., Quandt, R., 1982. Racetrack betting and informed behavior. Journal of Financial Economics 10, 187–194.

Barghouty, P., Corbett-Davies, S., Jenson, D., Overgoor, J., Phillips, C., Pierson, E., Ramachandran, V., Shoemaker, A., Shroff, R., Simoiu, C., 2020. A large-scale analysis of racial disparities in police stops across the United States. Nature Human Behaviour 4, 736–745.

Bayer, P., Ferreira, F., Ross, S., 2018. What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders. Review of Financial Studies 31, 175–205.

Becker, G.S., 1971. The economics of discrimination. University of Chicago Press.

Bertrand, M., Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American Economic Review 94, 991–1013.

Binder, A., Grimes, P., 2022. Defying the odds: Horseplayers and female jockeys. Applied Economics Letters 29, 1556–1563.

Binder, A., Grimes, P., Winterbotham, R., 2021. In the money: Gender and jockey success on the thoroughbred racetrack. Journal of Sports Economics 22, 295–328.

Brown, A., Yang, F., 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., Mayer, E., Weston, J., 2023. Racial disparities in the auto loan market. Review of Financial Studies 36, 1–41.

Camerer, C., 1998. Can asset markets be manipulated? A field experiment with racetrack betting. Journal of Political Economy 106, 457–482.

Cashmore, V., Coster, N., Forrest, D., McHale, I., Buraimo, B., 2022. Female jockeys – what are the odds? Journal of Economic Behavior & Organization 202, 703–713.

Chiappori, P., Gandhi, A., Salanié, B., Salanié, F., 2019. From aggregate betting data to individual risk preferences. Econometrica 87, 1–36.

Crafts, N., 1985. Some evidence of insider knowledge in horse race betting in Britain. Economica 52, 295–304.

Cronley, J., 2000. The not-so-glorious life of show betting. ESPN. URL Link.

Doleac, J., Stein, L., 2013. The visible hand: Race and online market outcomes. Economic Journal 123, F469–F492.

Dougal, C., Gao, P., Mayew, W., Parsons, C., 2019. What's in a (school) name? Racial discrimination in higher education bond markets. Journal of Financial Economics 134, 570–590.

Eggers, A., Freier, R., Grembi, V., Nannicini, T., 2018. Regression discontinuity designs based on population thresholds: Pitfalls and solutions. American Journal of Political Science 62, 210–229.

Ewens, M., Townsend, R., 2020. Are early stage investors biased against women? Journal of Financial Economics 135, 653–677.

Feess, E., Müller, H., Schumacher, C., 2014. The favorite–longshot bias and the impact of experience. Business Research 7, 217–234.

Feigenberg, B., Miller, C., 2022. Would eliminating racial disparities in motor vehicle searches have efficiency costs? Quarterly Journal of Economics 137, 49–113.

Fenly, B., 2023. Purses set record in 2022, but handle drops slightly. Thoroughbred Daily News. URL Link.

Figlewski, S., 1979. Subjective information and market efficiency in a betting market. Journal of Political Economy 87, 75–88.

Forrest, D., Kainulainen, T., Saastamoinen, J., Suhonen, N., 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., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. Predictably unequal? The effects of machine learning on credit markets. Journal of Finance 77, 5–47.

Gandar, J.M., Zuber, R.A., Johnson, R.S., 2001. Searching for the favourite-longshot bias down under: An examination of the New Zealand pari-mutuel betting market. Applied Economics 33, 1621–1629.

Gandhi, A., Serrano-Padial, R., 2015. Does belief heterogeneity explain asset prices: The case of the longshot bias. Review of Economic Studies 82, 156–186.

Geertsema, P., Schumacher, C.R., 2016. Parimutuel contests with strategic risk-sensitive bettors. Applied Economics 48, 1140–1158.

Golec, J., Tamarkin, M., 1998. Bettors love skewness, not risk, at the horse track. Journal of Political Economy 106, 205–225.

Gornall, W., Strebulaev, I.A., 2024. Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels. Management Science .

Green, E., Lee, H., Rothschild, D., 2020. The favorite–longshot Midas. SSRN 3271248.

Griffith, R., 1949. Odds adjustments by American horse-race bettors. American Journal of Psychology 62, 290–294.

Grumbach, J., Sahn, A., 2020. Race and representation in campaign finance. American Political Science Review 114, 206–221.

Guell, M., Rodriguez Mora, J., Telmer, C., 2015. The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. Review of Economic Studies 82, 693–735.

Han, L., Huang, X., Kadan, O., Wu, J., 2022. The true colors of money: Racial diversity and asset management. SSRN 3966470.

Hausch, D., Rubinstein, M., Ziemba, W., 1981. Efficiency of the market for racetrack betting. Management Science 27, 1435–1452.

Hausch, D., Ziemba, W., 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.

HKJC, 2023. The Hong Kong Jockey Club 2023 Annual Report. URL Link.

Hundtofte, S., Meyer, A., 2023. The longshot bias is a context effect. Management Science 69, 6954–6968.

Hurley, W., McDonough, L., 1995. A note on the Hayek hypothesis and the favorite–longshot bias in parimutuel betting. American Economic Review 85, 949–955.

Igan, D., Pinheiro, M., Smith, J., 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., Lee, Y., 2022. Racial disparities and potential channels in the NBA digital collectibles market. SSRN 4236789.

Kumar, A., Niessen-Ruenzi, A., Spalt, O., 2015. What's in a name? Mutual fund flows when managers have foreign-sounding names. Review of Financial Studies 28, 2281–2321.

Larsen, T., Price, J., Wolfers, J., 2008. Racial bias in the NBA: Implications in betting markets. Journal of Quantitative Analysis in Sports 4, 1–19.

Law, D., Peel, D., 2002. Insider trading, herding behaviour and market plungers in the British horse race betting market. Economica 69, 327–338.

Leeds, M., Rockoff, H., 2019. Beating the odds: Black jockeys in the Kentucky Derby 1870–1911. Historical Perspectives on Sports Economics 1, 136–149.

Leeds, M., Rockoff, H., 2020. Jim Crow in the saddle: The expulsion of African American jockeys from American racing. National Bureau of Economic Research 28167, 1–58.

Lu, Y., Naik, N.Y., Teo, M., 2024. Race, discrimination, and hedge funds. SSRN 4070484.

Moskowitz, T., 2021. Asset pricing and sports betting. Journal of Finance 76, 3153–3209.

Moskowitz, T., Vasudevan, K., 2022. Betting without beta. SSRN 3845505.

Nelson, A., 2020. What is a show bet in horse racing? Offshore Sports. URL Link.

Niessen-Ruenzi, A., Ruenzi, S., 2019. Sex matters: Gender bias in the mutual fund industry. Management Science 65, 3001–3025.

Ottaviani, M., Sørensen, P., 2008. The favorite–longshot bias: An overview of the main explanations. Handbook of Sports and Lottery Markets 1, 83–101.

Ottaviani, M., Sørensen, P., 2010. Noise, information, and the favorite–longshot bias in parimutuel predictions. American Economic Journal: Microeconomics 2, 58–85.

Price, J., Wolfers, J., 2010. Racial discrimination among NBA referees. Quarterly Journal of Economics 125, 1859–1887.

Purdum, D., 2024. Record $210.7M bet on 2024 Kentucky Derby. ESPN. URL Link.

Quandt, R., 1986. Betting and equilibrium. Quarterly Journal of Economics 101, 201–207.

Roeder, O., 2023. I used AI to bet on horse-racing. Here's what happened. Financial Times. URL Link.

Schnytzer, A., Shilony, Y., 1995. Inside information in a betting market. Economic Journal 105, 963–971.

Snowberg, E., Wolfers, J., 2010. Explaining the favorite–longshot bias: Is it risk-love or misperceptions? Journal of Political Economy 118, 723–746.

Snyder, W., 1978. Horse racing: Testing the efficient markets model. Journal of Finance 33, 1109–1118.

Thaler, R., Ziemba, W., 1988. Anomalies: Parimutuel betting markets: Racetracks and lotteries. Journal of Economic Perspectives 2, 161–174.

U.K. Gambling Commission, 2024. Industry Statistics - February 2024 - Correction. URL Link.

Vandenbruaene, J., Ceuster, M.D., Annaert, J., 2022. Efficient spread betting markets: A literature review. Journal of Sports Economics 23, 907–949.

Weitzman, M., 1965. Utility analysis and group behavior: An empirical study. Journal of Political Economy 73, 18–26.

Yakowicz, W., 2023. U.S. set gambling record in 2022 with more than $54.9 billion in revenue. Forbes. URL Link.

Ziemba, W., 2019. Exotic betting at the racetrack. World Scientific.

Figures and Tables

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 Appendix Figure A.1.

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	Accidental Hero (KY)		3/C	L	$70,000	F Arrieta	122	S M Asmussen	12/1
2	2	Judo (KY)		3/C	L	$80,000	R Santana, Jr.	125	B H Cox	3/1
3	3	Giroovin (FL)		3/C	L	$80,000	I Castillo	122	T W Fincher	7/2
4	4	Sir (KY)		3/C	L	$50,000	W De La Cruz	116	C Contreras	6/1
5	5	Tres Soles (KY)		3/C	L	$70,000	C A Torres	119	S M Asmussen	8/1
6	6	Recker Point (KY)		3/C	L	$80,000	M Murrill	125	C A Hartman	8/5
7	7	G T Five Hundred (KY)		3/G	L	$80,000	T Wales	122	I Mason	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 2: Trainer Surname Categorization

The figure graphs a scatter plot of 2,634 trainer surnames that appear in the U.S. horse racing logs and the Census Frequently Occurring Surnames file. 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 surname categorization. Of 2,915 total trainers, 1,608 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). The data are from 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1.

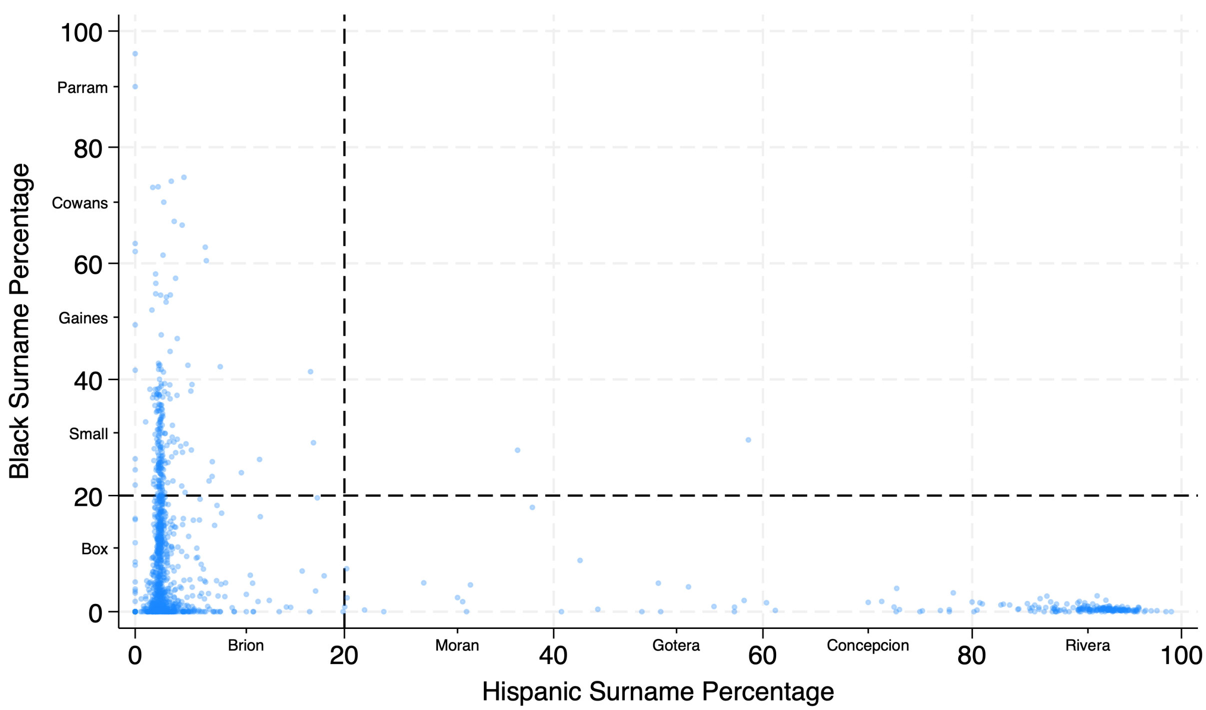


Figure 3: Realized Returns by Odds

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 2011–2022 U.S. horse racing logs as described in Section 3.1. Variable definitions are provided in Appendix Table A.1.

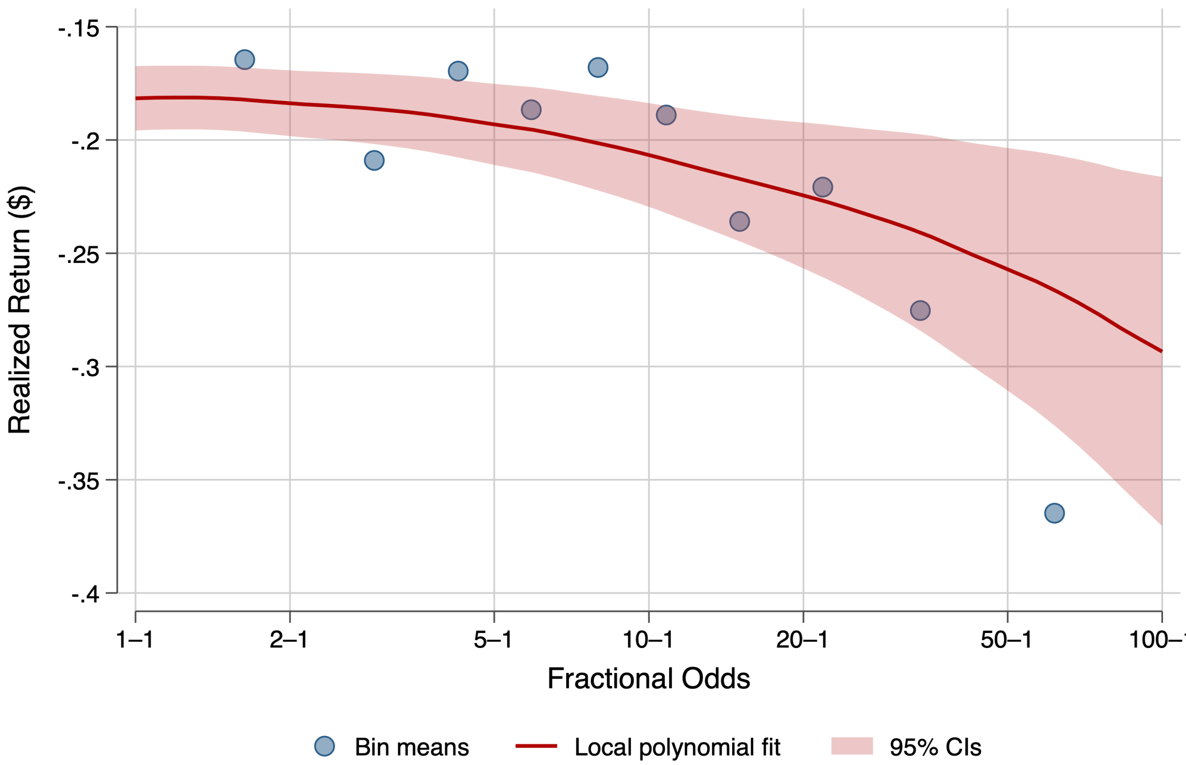


Figure 4: Realized Returns by Trainer Surname Category

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 Appendix Table A.1.

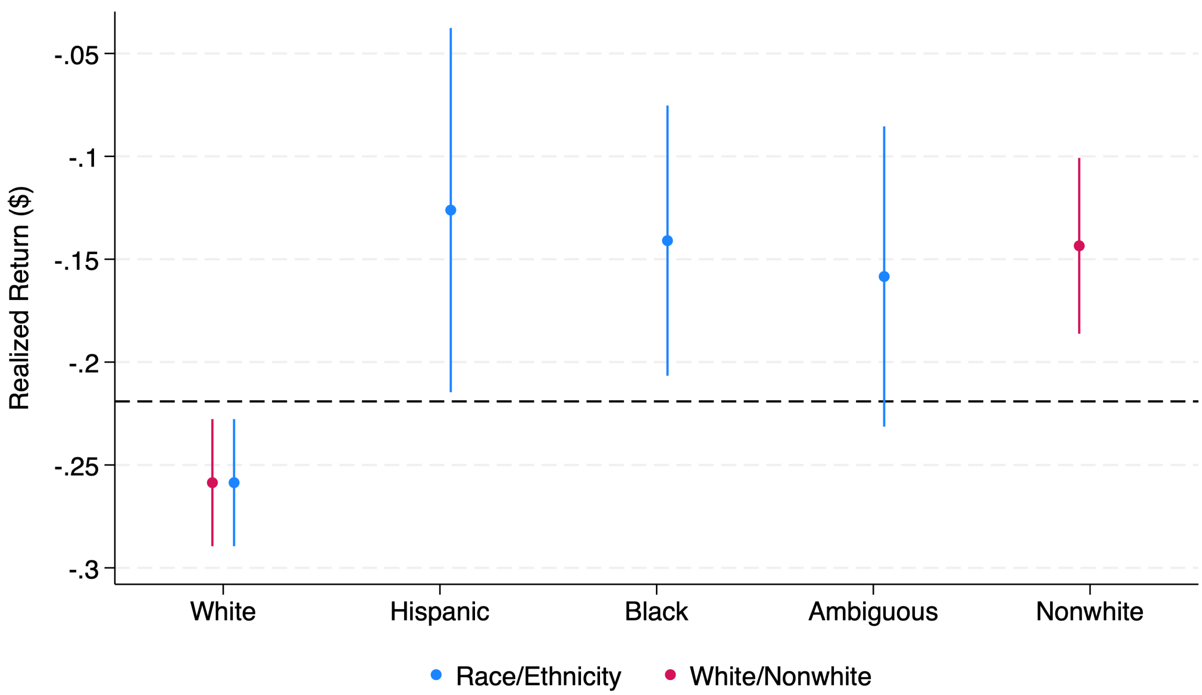


Figure 5: Odds Distribution by White/Nonwhite Trainer Surname

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 Appendix Table A.1.

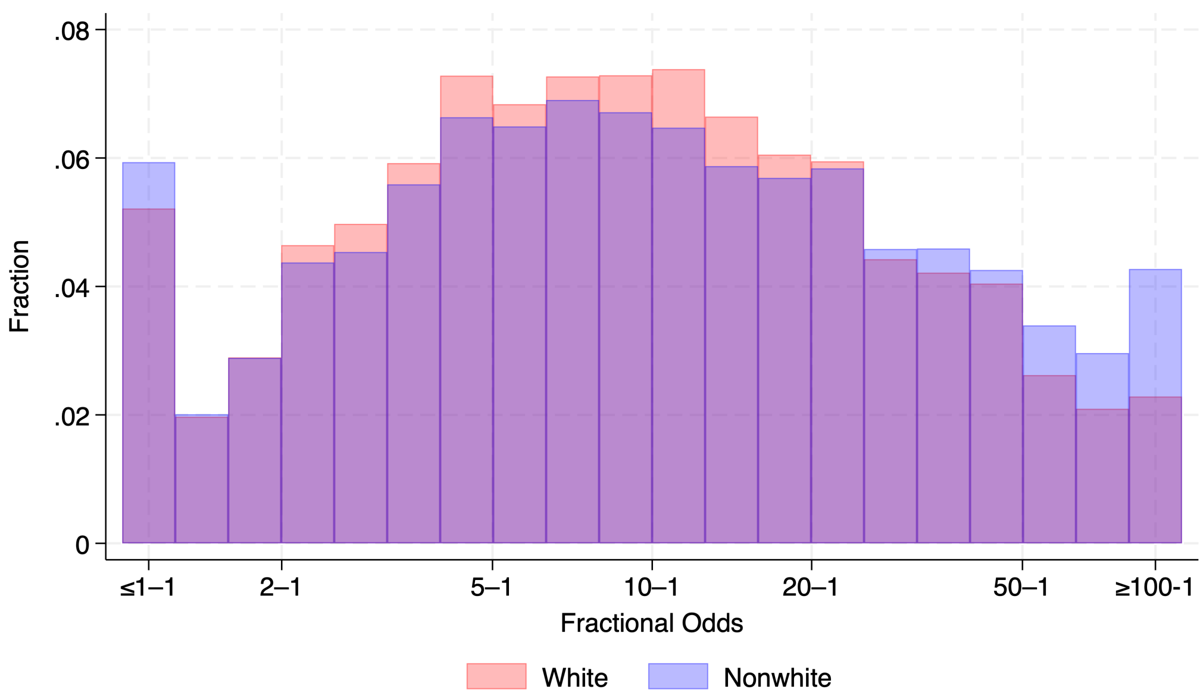


Figure 6: Realized Returns by Odds and White/Nonwhite Trainer Surname

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 nonwhite trainer surnames. 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 Appendix Table A.1.

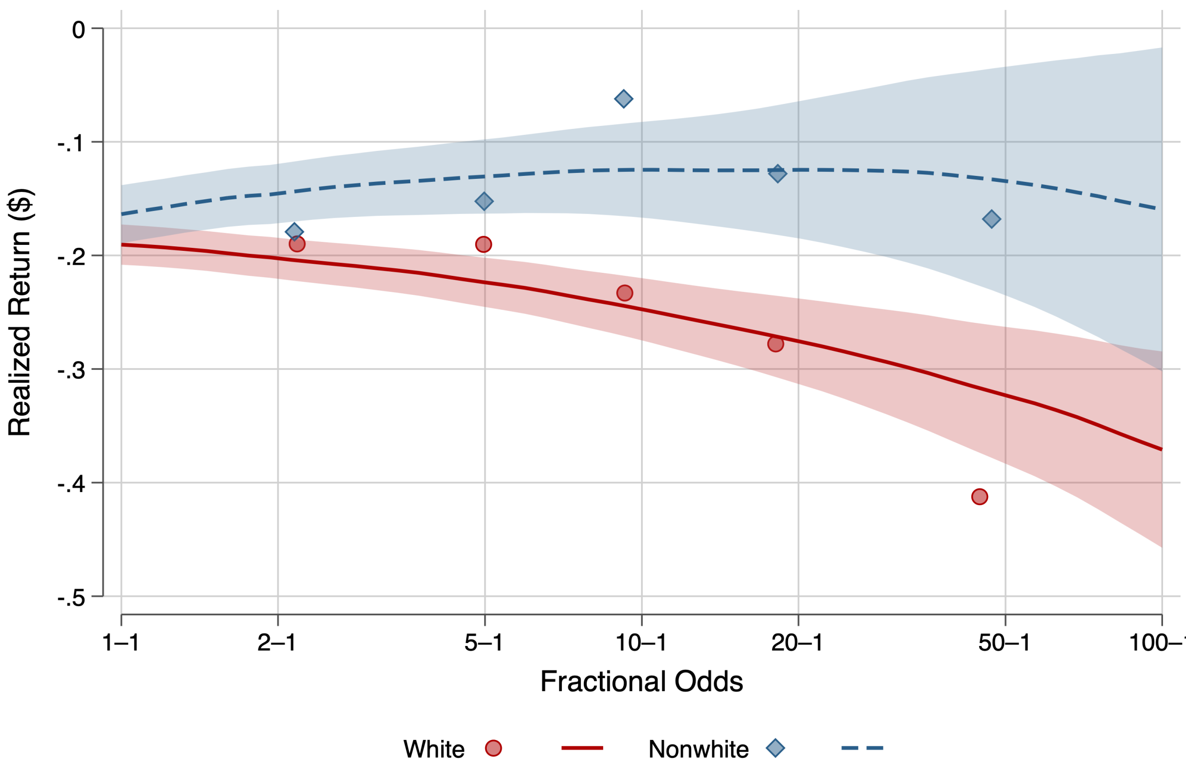


Table 1: Summary Statistics

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 2011–2022 U.S. horse racing logs and the U.S. Census Bureau as described in Section 3.1. Variable definitions are provided in Appendix Table A.1.

	(1) Mean	(2) SD	(3) Median
Horserace/Bet Outcomes			
Realized Return ($)	-0.22	3.50	–1.00
Position	4.90	2.88	5.00
Won	0.12		
Odds			
Decimal Odds	0.15	0.15	0.10
Fractional Odds	17.60	22.88	8.90
Fractional Odds (log)	2.19	1.22	2.19
Odds-Predicted Position	4.93	2.89	5.00
Trainer Surname Category			
White	0.65		
Nonwhite	0.35		
Hispanic	0.08		
Black	0.15		
Ambiguous	0.12		
Horserace Attributes			
Runners	8.85	2.40	9.00
Purse ($)	181,894	359,531	96,153

Table 2: Realized Returns by Odds

The table presents results from the estimation of equation 1. The dependent variable is the realized return from a $1 win bet. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, respectively. The data are from 2011–2022 U.S. horse racing logs as described in Section 3.1. Variable definitions are provided in Appendix Table A.1.

	Dependent Variable: Realized Return ($)				
	(1)	(2)	(3)	(4)	(5)
Constant	-0.2191***	-0.1166***	-0.1066***	-0.1649***	-0.1342***
	(0.0128)	(0.0200)	(0.0238)	(0.0111)	(0.0198)
Odds (log)		-0.0468***	-0.0613***	-0.0131	-0.0369**
		(0.0126)	(0.0194)	(0.0111)	(0.0162)
Non-Favorite			0.0437	0.2887**	0.2353*
			(0.0376)	(0.1208)	(0.1253)
Non-Favorite × Odds				-0.1068**	-0.0765*
				(0.0423)	(0.0454)
Horserace FE					✓
N	74,988	74,988	74,988	74,988	74,988

Table 3: Horserace/Bet Outcomes by Odds and Trainer Surname Category

The table presents results from the 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 finishes are dropped from the sample). The dependent variable in columns (5) and (6) is a binary variable that takes the value of one if a horse won the horserace and zero otherwise; the resulting estimates therefore reflect a Linear Probability Model. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	(1)	(2)	(3)	(4)	(5)	(6)
	Realized Return		Position		Won	
Constant	-0.1542***	-0.1522***	2.2569***	2.2554***	-0.0074***	-0.0075***
	(0.0235)	(0.0235)	(0.0186)	(0.0186)	(0.0016)	(0.0016)
Nonwhite	0.1190***		-0.0622***		0.0081***	
	(0.0285)		(0.0185)		(0.0023)	
Hispanic		0.1534***		-0.1011***		0.0088**
		(0.0591)		(0.0324)		(0.0038)
Black		0.1186***		-0.0466*		0.0105***
		(0.0399)		(0.0256)		(0.0033)
Ambiguous		0.0962**		-0.0552**		0.0046
		(0.0421)		(0.0276)		(0.0036)
Odds (log)	-0.0482***	-0.0492***				
	(0.0125)	(0.0126)				
Predicted Position			0.5414***	0.5417***		
			(0.0034)	(0.0034)		
Odds (decimal)					0.8513***	0.8517***
					(0.0104)	(0.0104)
N	74,988	74,988	74,493	74493	74,988	74,988

Table 4: Realized Returns by Odds and White/Nonwhite Trainer Surname

The table presents results from the estimation of variations of equation 3. The dependent variable is the realized return from a $1 win bet. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

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

Table 5: Win, Place, and Show Realized Returns by Odds and White/Nonwhite Trainer Surname

The table presents results from the 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, respectively. 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 Appendix Table A.1.

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

Table 6: Realized Returns and Prior Finishing Position (Momentum)

The table presents results from the 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 sample of horses that have previous horserace observations. Columns (3) and (4) show the sub-sample of horses that finished in the last half of the field in their previous horserace. Columns (5) and (6) contain the sub-sample of horses that finished in the top half of the field in their previous horserace. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		Prior Finish Bottom Half		Prior Finish Top Half	
Constant	-0.1899***	-0.1622***	-0.1359**	-0.0840	-0.2038***	-0.1943***
	(0.0341)	(0.0312)	(0.0672)	(0.0606)	(0.0382)	(0.0371)
Nonwhite	0.1072***	0.0336	0.1650***	0.0274	0.0636	0.0377
	(0.0361)	(0.0248)	(0.0622)	(0.0460)	(0.0410)	(0.0293)
Odds (log)	-0.0217	-0.0245	-0.0545	-0.0594	-0.0091	-0.0098
	(0.0252)	(0.0251)	(0.0429)	(0.0424)	(0.0322)	(0.0323)
Non-Favorite	0.0132	-0.0356	0.0583	-0.0118	-0.0136	-0.0363
	(0.0463)	(0.0528)	(0.0771)	(0.0874)	(0.0576)	(0.0654)
Nonwhite × Non-Favorite		0.1613**		0.2358**		0.0729
		(0.0770)		(0.1110)		(0.1067)
N	47,165	47,165	20,715	20,715	26,450	26,450

Table 7: Realized Returns and Purse Size

The table presents results from the 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 with a purse size below below the average. Columns (5) and (6) contain the sub-sample of horses that run in horseraces with a purse size above the average. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		Small Purse		Large Purse	
Constant	-0.1438***	-0.1146***	-0.1192***	-0.0804***	-0.1684***	-0.1483***
	(0.0269)	(0.0248)	(0.0319)	(0.0299)	(0.0436)	(0.0396)
Nonwhite	0.1193***	0.0423**	0.1407***	0.0398	0.1010**	0.0475
	(0.0285)	(0.0208)	(0.0374)	(0.0285)	(0.0435)	(0.0304)
Odds (log)	-0.0634***	-0.0663***	-0.0941***	-0.0975***	-0.0340	-0.0362
	(0.0193)	(0.0192)	(0.0231)	(0.0230)	(0.0308)	(0.0305)
Non-Favorite	0.0459	-0.0010	0.0928*	0.0248	-0.0006	-0.0300
	(0.0376)	(0.0420)	(0.0504)	(0.0558)	(0.0556)	(0.0626)
Nonwhite $\times$ Non-Favorite		0.1532***		0.2030***		0.1060
		(0.0565)		(0.0750)		(0.0855)
N	74,988	74,988	37,516	37,516	37,472	37,472

Table 8: Realized Returns and Track Conditions

The table presents results from the 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 with fast track conditions. Columns (5) and (6) contain the sub-sample of horses that run in horseraces with non-fast track conditions. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		Fast Track Conditions		Other Track Conditions	
Constant	-0.1438***	-0.1146***	-0.1697***	-0.1320***	-0.1062***	-0.0877**
	(0.0269)	(0.0248)	(0.0368)	(0.0337)	(0.0387)	(0.0358)
Nonwhite	0.1193***	0.0423**	0.1142***	0.0159	0.1244***	0.0750**
	(0.0285)	(0.0208)	(0.0401)	(0.0274)	(0.0402)	(0.0317)
Odds (log)	-0.0634***	-0.0663***	-0.0520*	-0.0554**	-0.0812***	-0.0833***
	(0.0193)	(0.0192)	(0.0274)	(0.0273)	(0.0266)	(0.0264)
Non-Favorite	0.0459	-0.0010	0.0669	0.0031	0.0316	0.0036
	(0.0376)	(0.0420)	(0.0530)	(0.0597)	(0.0528)	(0.0586)
Nonwhite × Non-Favorite		0.1532***		0.1990**		0.0964
		(0.0565)		(0.0807)		(0.0790)
N	74,988	74,988	39,254	39,254	35,734	35,734

Table 9: Realized Returns in the U.S. South

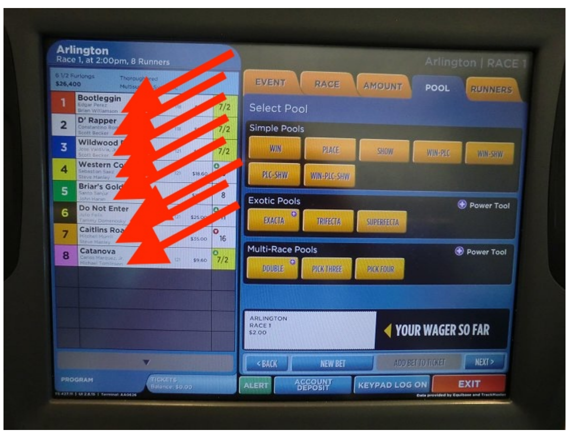
The table presents results from the estimation of variations of equation 3. The dependent variable is the realized return from a \$1 win bet. Columns (1) and (2) report pooled regressions using the full sample of observations. Columns (3) and (4) show the sub-sample of horses that run in horseraces at 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

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

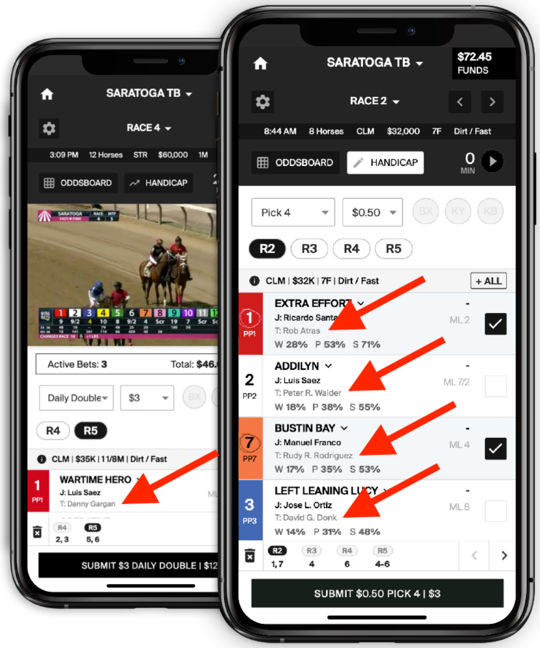
Appendix

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

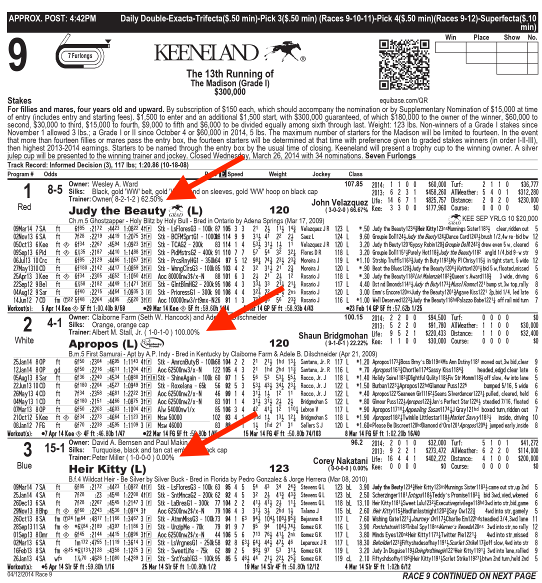
The figure illustrates various sources of information available to bettors, with the prominently featured trainer names indicated with red arrows.



(a) *Terminal Bet*



(b) *Phone Bet*



(c) *Paper Race Card*

Figure A.2: 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 6), 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 Appendix Table A.1.

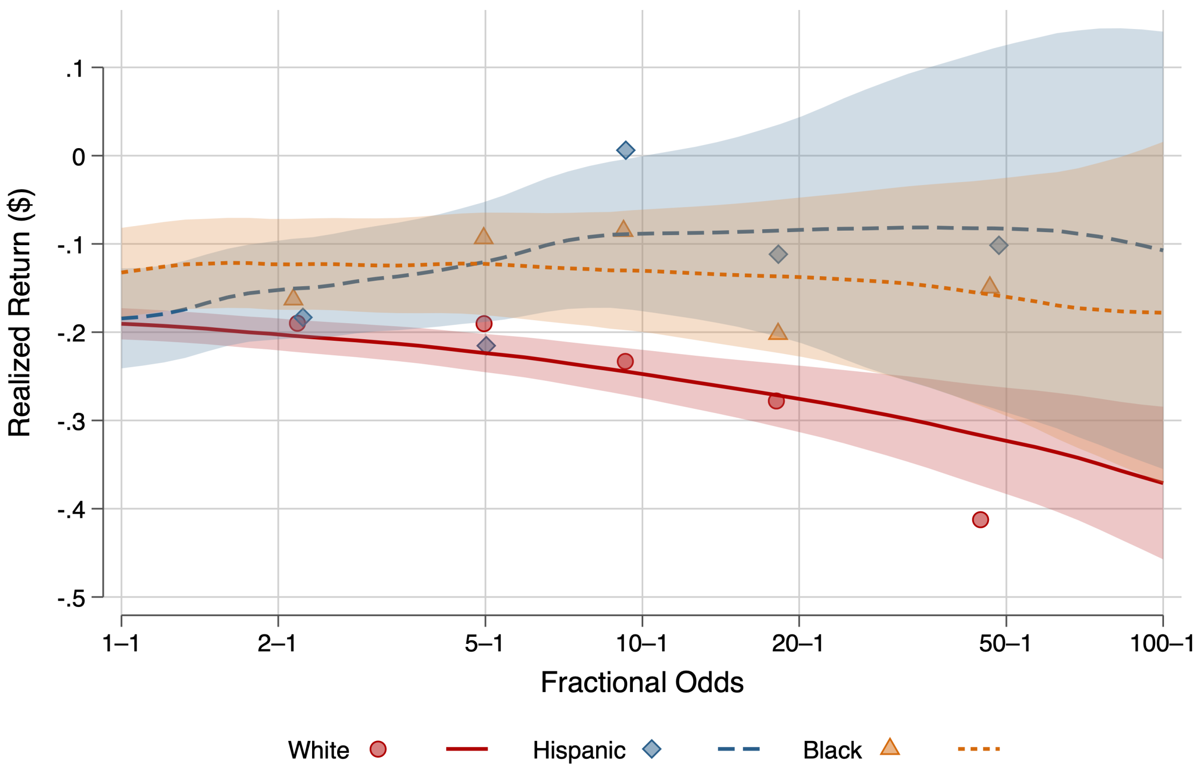


Table A.1: Variable Definitions

Variable	Definition
Horserace/Bet Outcomes	
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 one if a horse wins a race and zero otherwise.
Odds	
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 (for example, 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.
Non-Favorite	A binary indicator variable that takes the value of one if fractional odds are greater than the sample median, and zero otherwise. We consider alternate definitions in Appendix Table A.8.
Trainer Surname Category	
Nonwhite	A binary indicator variable that takes the value of one 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 zero otherwise. We consider alternate definitions in Appendix Table A.7.
Hispanic	A binary indicator variable that takes the value of one if a trainer's surname is greater than 20% Hispanic and less than 20% Black; and zero otherwise.
Black	A binary indicator variable that takes the value of one if a trainer's surname is greater than 20% Black and less than 20% Hispanic; and zero otherwise.
Ambiguous	A binary indicator variable that takes the value of one 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 zero otherwise.
Horserace Attributes	
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: 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

Table A.3: Realized Returns by Odds and Trainer Surname Race/Ethnicity

The table presents results from the estimation of variations of equation 3. The dependent variable is the realized return from a $1 win bet. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

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

Table A.4: Outcomes by Morning Line Odds and Trainer Surname Category

The table presents results from the 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 is 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 finishes are dropped from the sample), and a binary variable equal to one 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, respectively. 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 Appendix Table A.1.

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

Table A.5: Realized Returns by Odds and White/Nonwhite Jockey Surname

The table presents results from the estimation of variations of equation 3 using jockey surnames. There are 1,272 unique jockey names in the sample. 353 jockeys have white surnames, while 919 have nonwhite surnames. Out of the nonwhite 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

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

Table A.6: Realized Returns and Within-Horserace Trainer Surname Diversity

The table presents results from the 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 with a nonwhite average greater than the average of nonwhite. Columns (5) and (6) contain the sub-sample of horses that run in horseraces with a nonwhite average less than the average of nonwhite. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		More Diverse		Less Diverse	
Constant	-0.1438***	-0.1146***	-0.1154***	-0.0977***	-0.1771***	-0.1426***
	(0.0269)	(0.0248)	(0.0334)	(0.0300)	(0.0427)	(0.0412)
Nonwhite	0.1193***	0.0423**	0.1224***	0.0455	0.1198***	0.0570**
	(0.0285)	(0.0208)	(0.0473)	(0.0337)	(0.0396)	(0.0288)
Odds (log)	-0.0634***	-0.0663***	-0.0693***	-0.0712***	-0.0582*	-0.0609*
	(0.0193)	(0.0192)	(0.0238)	(0.0235)	(0.0310)	(0.0311)
Non-Favorite	0.0459	-0.0010	0.0169	-0.0107	0.0845	0.0256
	(0.0376)	(0.0420)	(0.0482)	(0.0533)	(0.0589)	(0.0671)
Nonwhite × Non-Favorite		0.1532***		0.1494		0.1269
		(0.0565)		(0.0923)		(0.0800)
N	74,988	74,988	41,238	41,238	33,750	33,750

Table A.7: Alternative Measures of Nonwhite Trainer Surnames

The table presents results from the 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 nonwhite surnames. In Panel (b), nonwhite 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

(a) Binary Nonwhite Categorization Using 80% Thresholds

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

(b) Continuous (Standardized) Nonwhite Measure

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

Table A.8: Alternative Measures of Favorite vs. Non-Favorite

The table presents results from the estimation of variations of equation 3 using different measures of odd cuts. Columns (1) and (2) show the non-favorite cut at the median (which is the baseline from Table 4). Columns (3) and (4) use the within-horserace median (i.e., half the horses are considered favorites). Columns (5) and (6) show the non-favorite cut at the 25th percentile. Columns (7) and (8) show the non-favorite cut using a quintile split where the base relation is the first quintile. The dependent variable is the realized return from a $1 win bet. Coefficient estimates are shown, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$p50$ Split (Baseline)		$p50$ Split (w/i horserace)		$p25$ Split		Quintile Split	
Constant	-0.1438***	-0.1146***	-0.1441***	-0.1140***	-0.1765***	-0.1368***	-0.1815***	-0.1414***
	(0.0269)	(0.0248)	(0.0246)	(0.0226)	(0.0198)	(0.0179)	(0.0219)	(0.0206)
Nonwhite	0.1193***	0.0423**	0.1191***	0.0371*	0.1214***	0.0082	0.1222***	0.0146
	(0.0285)	(0.0208)	(0.0285)	(0.0221)	(0.0284)	(0.0211)	(0.0283)	(0.0221)
Odds (log)	-0.0634***	-0.0663***	-0.0714***	-0.0738***	-0.0826***	-0.0843***	-0.0725**	-0.0791**
	(0.0193)	(0.0192)	(0.0168)	(0.0167)	(0.0203)	(0.0202)	(0.0339)	(0.0337)
Non-Favorite	0.0459	-0.0010	0.0770**	0.0288	0.1298***	0.0815**		
	(0.0376)	(0.0420)	(0.0307)	(0.0356)	(0.0366)	(0.0400)		
Nonwhite × Non-Favorite		0.1532***		0.1533***		0.1519***		
		(0.0565)		(0.0541)		(0.0428)		
Nonwhite × Odds q2								0.0141
								(0.0415)
Nonwhite × Odds q3								0.1397**
								(0.0543)
Nonwhite × Odds q4								0.1410**
								(0.0718)
Nonwhite × Odds q5								0.2407**
								(0.1055)
Odds Quintile FE							✓	✓
N	74,988	74,988	74,988	74,988	74,988	74,988	74,988	74,988

Table A.9: Controlling for Perceived Trainer Quality (Across- and Within-Trainer)

The table presents results from the 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

	Dependent Variable: Realized Return ($)			
	(1)	(2)	(3)	(4)
Constant	-0.1127***	-0.1227***	-1.9188***	-1.7313***
	(0.0265)	(0.0265)	(0.0499)	(0.0467)
Nonwhite	0.0420**	0.0423**	-0.0856***	-
	(0.0210)	(0.0208)	(0.0188)	
Odds (log)	-0.0586***	-0.0651***	0.3825***	0.3438***
	(0.0199)	(0.0192)	(0.0239)	(0.0237)
Non-Favorite	-0.0220	-0.0011	-0.1246***	-0.1435***
	(0.0430)	(0.0420)	(0.0376)	(0.0382)
Nonwhite × Non-Favorite	0.1715***	0.1543***	0.2269***	0.1679***
	(0.0588)	(0.0565)	(0.0512)	(0.0553)
Trainer Won Last Race	-0.0771**			-1.0372***
	(0.0345)			(0.0383)
Number of Races in a Year for a Trainer		0.0002		-0.0004
		(0.0003)		(0.0004)
Trainer Win Percentage for a Track-Year			0.0716***	0.0735***
			(0.0014)	(0.0014)
Trainer FE				✓
N	72,073	74,988	74,988	71,675

Table A.10: Realized Returns by Odds and Trainer Surname Race/Ethnicity in the U.S. South

The table presents results from the 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, and their robust standard errors are displayed in parentheses. ***(**)(*) Significance at the 1%(5%)(10%) two tailed level, 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 Appendix Table A.1.

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