# Peer Effects in Horse Racing:
# A Tripartite Decomposition of Competitive Responses

Tandem Young

Department of Economics

University of Arkansas

April 16, 2026

**Abstract**

This paper estimates whether field composition causally affects individual horse performance using the 2023 Equibase dataset (42,618 races; 318,702 horse-race entries). A one-point increase in leave-out mean competitor quality raises a horse's own speed rating by 0.10–0.21 Beyer points across OLS, horse fixed effects, two independent instrumental variable strategies, and a claiming race subsample, with four of the five strategies concentrated in the 0.14–0.21 range. The paper's main contribution is a tripartite decomposition of peer effects into horse, jockey, and trainer channels: horses exhibit a positive competitive response to stronger fields (+0.188), while jockeys and trainers show negative tactical and strategic effects (−0.085 and −0.063, respectively). Better horses respond more strongly to competition—a convex, "rising to competition" pattern that contrasts with the discouragement effect documented in golf. All results survive Oster bounds, permutation tests, Hausman specification tests, causal forests, and collinearity diagnostics.

**JEL Codes:** D82, J24, L83, Z20

**Keywords:** peer effects, tournament theory, horse racing, tripartite decomposition, instrumental variables

# 1 Introduction

Does the quality of one's competitors causally affect individual performance? This question is central to tournament theory (Lazear and Rosen, 1981) and contest design, yet credible empirical answers remain scarce because competitor assignment is rarely exogenous. Manski (1993) formalizes the core difficulty as the *reflection problem*—the impossibility of separately identifying causal peer effects from selection into groups and from shared environmental shocks using only observational data. Thoroughbred horse racing offers an unusually clean setting to address this question: because competitors in the same contest simultaneously influence one another, disentangling the causal effect of peer quality from correlated shocks and strategic sorting has proven exceptionally difficult in most domains, but racing's institutional features generate quasi-random variation in field composition.

This paper exploits this natural laboratory for studying peer effects in competitive settings. Horse racing offers four features that, taken together, provide an unusually clean identification environment. First, individual output is standardized and continuous: Beyer speed figures translate finishing times into a track- and distance-adjusted performance metric, eliminating the measurement imprecision that plagues many studies of competitive output. Second, peer groups vary across observations for the same individual—the same horse faces different competitors from race to race—generating the within-unit variation in peer composition that identification requires. Third, racing is a multi-agent production process in which the contributions of horse, jockey, and trainer are separately identifiable, because the same horse races under different jockeys and trainers over its career; this structure permits a decomposition of aggregate peer effects into distinct channels that is infeasible in most settings. Fourth, institutional features of the racing industry generate quasi-exogenous variation in field composition: late veterinary scratches randomly alter the set of competitors, and claiming price constraints restrict entry to horses of similar assessed value. A date-based preference system for oversubscribed races provides further quasi-random variation in the background, though we do not exploit it as a formal instrument. These features provide multiple independent sources of identification, addressing the concern that any single instrument may be inadequate to resolve the reflection problem.

We estimate the causal effect of field composition on individual horse performance using the complete universe of 2023 North American thoroughbred races from Equibase, comprising 42,618 races and 318,702 individual entries. Across five identification strategies—pooled ordinary least squares (OLS) with extensive controls, horse fixed effects (FE), a scratch-based instrumental variable (IV), a network IV exploiting the partially overlapping structure of racing schedules, and a claiming race subsample where institutional constraints limit strategic entry—we find positive peer effect estimates ranging from 0.10 to 0.21 Beyer points, with four of the five strategies concentrated in the 0.14–0.21 range. The network instrumental variable yields a modestly attenuated estimate of 0.10, consistent with its identification of a different local average treatment effect. Sign consistency across all five approaches—each grounded in different identifying assumptions—is the central diagnostic (see Figure 1). In economic terms, a one-standard-deviation increase in field quality (approximately 14.8 Beyer points) translates to roughly 1.2 horse lengths at a typical distance—a meaningful margin in competitive racing, where the median winning margin is under two lengths.

The paper's principal contribution is a tripartite decomposition of peer effects into horse, jockey, and trainer channels. Using triple fixed effects and channel-specific peer quality measures, we show that horses exhibit a positive competitive response to stronger fields (+0.188 Beyer points per unit of peer quality), consistent with real-time pacing and competitive facilitation. Jockeys and trainers, by contrast, display negative tactical and strategic responses (−0.085 and −0.063, respectively), consistent with effort reallocation or conservative positioning when facing superior competition. The headline estimate of approximately 0.15 is therefore conservative: it nets a larger positive horse response against negative agent channels. Furthermore, we document substantial heterogeneity across the ability distribution. The strongest horses respond most positively to stronger fields—a convex, "rising to competition" pattern that contrasts sharply with the discouragement effect documented in professional golf by Brown (2011) and aligns with the nonlinear peer effects framework developed by Boucher et al. (2024).

These findings contribute to several strands of the literature. Relative to the influential discouragement result in Brown (2011), we find the opposite: competition facilitates rather than depresses performance, particularly among the most able. Relative to the null result in Guryan et al. (2009), who find no peer effects from random playing-group assignment in golf, our setting differs in ways that theory predicts should generate stronger peer influence—sustained physical proximity, real-time tactical interaction, and direct competition over a common distance. Building on Ashby (2023a), who exploits institutional variation in field composition in swimming, and Brox and Goller (2025), who documents convex responses in professional darts, we add the tripartite agent decomposition and a richer identification battery. To our knowledge, no published study has estimated peer effects on individual horse performance, nor has any paper in the broader peer effects literature decomposed competitive spillovers into the distinct channels of the primary competitor, the on-field decision maker, and the strategic planner.

The remainder of the paper proceeds as follows. Section 2 provides institutional background on thoroughbred racing, focusing on the features that generate identifying variation. Section 3 reviews the relevant literatures on tournament theory, peer effects, and horse racing economics. Section 4 describes the data and key variable construction. Section 5 develops the empirical strategy, detailing each of the five identification approaches. Section 6 presents the main results and the tripartite decomposition. Section 7 reports extensive robustness checks, including sensitivity to functional form, alternative peer quality measures, and placebo tests. Section 8 concludes.

## 2 Institutional Background

### The Tripartite Agent Structure

A thoroughbred horse race is a single contest over a fixed distance on a dirt or turf surface, with fields of two to eighteen runners. Distances range from five to twelve furlongs (one furlong equals one-eighth of a mile). Each entry involves three nested agents: the horse, which provides the competitive performance response; the jockey, who makes real-time tactical decisions during the race; and the trainer, who makes strategic pre-race decisions including race selection, fitness preparation, and equipment choices.

Crucially, the mapping between these agents varies across starts. The same horse races under different jockeys over the course of a season, and the same jockey rides for many different trainers and on many different horses. This recombination generates the cross-classified variation that enables the tripartite decomposition central to this paper: by observing the same horse with different jockeys and the same jockey on different horses, we can separately identify the competitive response operating through each agent channel.

### Race Types and Field Composition

Four race types determine eligibility and constrain the degree of strategic sorting in field composition (see Forrest, 2023, for a survey). *Maiden races* are restricted to horses that have not yet won, producing fields with lower and more variable quality. *Claiming races*, which account for approximately sixty percent of all U.S. races—and a comparable share of our estimation sample, require that every entered horse be available for purchase at a posted claiming price. The claiming price functions as a quality cap: a trainer will not enter a horse worth $50,000 in a $10,000 claiming race because any other trainer could purchase it for far below its true value. This mechanism constrains the quality range within a race and substantially limits strategic sorting, making field composition closer to random assignment than in other race types. Claiming races therefore provide the cleanest identification setting in our analysis. *Allowance races* set eligibility based on conditions such as win records or cumulative earnings, affording trainers somewhat more discretion in race selection. *Stakes races* represent the highest competitive level and involve the greatest degree of strategic entry.

### Entry Decisions and the Identification Threat

Trainers choose which races to enter based on horse fitness, expected competition, and distance and surface preferences. This strategic entry creates the core identification challenge: trainers who know their horse is in peak fitness may systematically enter races against stronger competition, generating a spurious positive correlation between field quality and own performance that reflects sorting rather than a causal peer effect. This is the reflection problem formalized by Manski (1993)—simultaneous determination of individual and group outcomes makes it difficult to separate the causal effect of peers from the correlated unobservables that drive selection into the same contest. Any credible identification strategy must isolate variation in field composition that is orthogonal to the trainer's private information about horse fitness at entry.

### Scratches as Quasi-Exogenous Variation

After entries close, typically forty-eight to seventy-two hours before post time, horses may be removed from the field through scratches. Three types of scratches occur. A regulatory veterinarian orders veterinary scratches; under the Horseracing Integrity and Safety Act (HISA) Rule 2262,[1] the veterinarian holds "unconditional authority" to remove any horse deemed unfit to compete. Racing officials impose steward scratches for administrative reasons. Trainer or owner scratches are voluntary withdrawals that may reflect private information about horse fitness, making them potentially endogenous. Veterinary and steward scratches alter field composition after the entry decision has been made, and therefore change the competitive environment in ways that are plausibly independent of the remaining horses' fitness on race day. This post-entry variation in field quality—driven by the health or regulatory status of *other* horses—provides the *exclusion restriction* (the requirement that the instrument affect the outcome only through the endogenous treatment variable) for the scratch-based instrumental variable strategy developed in Section 5. Additional quasi-random variation arises from the preference system: when races are oversubscribed, a date-based priority rule allocates starting positions, and also-eligible horses fill scratch-created vacancies in drawn order—neither mechanism reflects trainer information about the remaining field.

### Performance Measurement

The outcome variable throughout this paper is the Beyer speed figure, a standardized performance metric that adjusts raw finishing times for track surface, distance, and prevailing conditions. A Beyer figure of 100 represents elite performance, 80 is strong, and 60 is roughly average. In our data, one Beyer point corresponds to approximately 0.54 horse lengths at the finish, so the peer effects estimated in this paper translate to physically meaningful shifts in competitive positioning. Because the adjustment procedure is designed to produce comparability across tracks and distances, Beyer figures provide a continuous outcome measure suitable for regression analysis without requiring additional normalization.

### Terminology

Several racing-specific terms appear throughout the analysis and merit brief definition.[2]

[1] Horseracing Integrity and Safety Authority, Racetrack Safety Rule 2262. Full regulations available at https://hisaus.org/regulations.

[2] *Post time* is the scheduled start time of a race. A *card* is the day's lineup of races at a given track. The *first point of call* is the first mid-race position marker, typically at the quarter-mile mark, used to classify each horse's running style. Horses are commonly categorized as *front-runners* (those that race at or near the lead from the start), *stalkers* (those that sit mid-pack and advance in the middle stages), or *closers* (those that trail early and accelerate in the final stages). The *pacesetter* is the horse setting the early pace. *Drafting* refers to running immediately behind

## 3 Literature Review

This paper sits at the intersection of several literatures: tournament theory and competitor interactions, peer effects in sports, the econometrics of peer effects identification, and the economics of horse racing. This section reviews each in turn, identifying the specific gaps that motivate our empirical approach.

### 3.1 Tournament Theory and Competitor Interactions

The theoretical foundation for analyzing competitive interactions begins with Lazear and Rosen (1981), who model rank-order tournaments as optimal labor contracts. In their framework, a principal offers prizes that depend on ordinal rank rather than absolute output, and workers optimally choose effort based on the *spread* between prizes rather than the prize level itself. The model generates sharp predictions: effort increases in the prize gap and decreases in the noise of the production process. Crucially, effort also depends on the composition of competitors, since the marginal return to effort varies with the ability distribution of one's rivals. This theoretical channel—that *who* you compete against affects *how* you perform—provides the core motivation for our empirical investigation.

The most influential empirical test of how competitor composition affects effort is Brown (2011), who documents a "superstar" or discouragement effect in professional golf. When Tiger Woods enters a PGA Tour event, non-superstar competitors score approximately 0.8 strokes worse, consistent with rational effort reduction when the probability of winning falls. Brown's finding has shaped the literature's default expectation that stronger competitors depress rivals' performance. Whether this discouragement effect generalizes to settings with different competitive structures—particularly those featuring simultaneity, physical proximity, and real-time interaction—remains an open question that motivates our investigation.

Hill (2014) offers a partial answer. Studying 100-meter sprint heats, Hill finds that runners post faster times in heats that include Usain Bolt—a positive superstar effect. He attributes this

another horse to benefit from reduced air resistance. The *weight carried* is the combined weight of the jockey and equipment assigned by the racing secretary as a handicap; heavier weight mechanically slows the horse. A *layoff* is the time elapsed since a horse's last start. *Morning line odds* are the track handicapper's pre-betting odds estimate, which serve as an ex-ante quality proxy in some analyses. *Also-eligible* horses are those on an alternate list who fill vacancies created by scratches, in the order drawn.

reversal to the simultaneous, non-strategic nature of sprinting: when competitors run side-by-side in a brief event, the physical presence of a fast rival serves as a pacing mechanism rather than a source of discouragement. Horse racing occupies an instructive middle ground between Brown's golf setting and Hill's sprinting context. Like sprinting, racing is simultaneous and relatively brief, creating scope for real-time facilitation. Yet unlike sprinting, races unfold over sufficient distance to permit positional strategy—stalking, drafting, and tactical acceleration—suggesting that strategic responses to competitor quality also operate.

Heterogeneous responses to competition have received growing attention. Boudreau et al. (2016) find nonmonotonic responses in software development contests: most contestants are discouraged by stronger fields, but the highest-ability participants respond positively. This heterogeneity motivates our examination of whether responses to field quality vary across the ability distribution. Brox and Goller (2025) provide further evidence in professional darts using causal forests, finding that contestant heterogeneity is detrimental for lower-ability players but beneficial for higher-ability ones—a methodological template for our own heterogeneity analysis.

Bilen and Matros (2023) decompose the superstar effect in chess into direct (head-to-head) and indirect (mere presence in the same tournament) channels, a distinction relevant to our separation of within-race peer effects from selection into races. A meta-analysis by Drugov and Ryvkin (2025) identifies conditions under which facilitation versus discouragement is more likely to emerge. Recent theoretical work by Boucher et al. (2024) develops a general framework for nonlinear peer effects that nests both concave (discouragement-consistent) and convex (facilitation-consistent) functional forms, motivating our attention to nonlinearity in estimation.

### 3.2 Peer Effects in Sports

A complementary literature estimates peer effects in athletic settings, exploiting the relatively clean measurement of individual performance that sports provide. The landmark study by Guryan et al. (2009) finds no evidence of peer effects from random playing-group assignment in professional golf: a golfer's score is unrelated to the ability of his randomly assigned playing partners. This null result is important both substantively and methodologically. Substantively, it suggests that in settings where competitors interact minimally during performance—four-hour golf rounds involve little direct strategic interaction—peer effects may be negligible. Methodologically, Guryan, Kroft, and Notowidigdo demonstrate the importance of controlling for exclusion bias in leave-out mean specifications, a concern we address following Caeyers and Fafchamps (2024). Our setting differs from golf in a critical respect: horse races involve sustained physical proximity, real-time positional maneuvering, and direct tactical interaction, all of which create stronger channels for peer influence.

Building on the random-assignment paradigm established by Sacerdote (2001), two recent studies of high school athletics are particularly relevant to our approach. Ashby (2023a) estimates peer effects using classification realignments in swimming and cross-country running, exploiting institutional variation in field composition that is plausibly orthogonal to individual ability trends. Ashby (2023b) extends this analysis to document heterogeneous peer effects in cross-country, finding that responses to peer quality vary across the ability distribution. These papers are the closest methodological analogs to our work in that they use institutional variation in competitor assignment to identify causal peer effects on individual athletic performance.

Similar facilitation dynamics have been documented in swimming, where no-shows in adjacent lanes slow remaining competitors (Yamane and Hayashi, 2015), and in marathon running, where designated pace setters improve elite finishing times (Emerson and Hill, 2018). Both findings support the real-time pacing mechanism we hypothesize operates in horse racing.

### 3.3 Identification of Peer Effects

The econometric challenge of identifying causal peer effects has been a central concern since Manski (1993) formalized the reflection problem. Manski demonstrated that in a linear-in-means model, one cannot separately identify endogenous peer effects (the causal effect of peers' outcomes), exogenous or contextual effects (the effect of peers' predetermined characteristics), and correlated effects (common shocks or sorting) without additional structure. This tripartite identification challenge—or some version of it—confronts every peer effects study, including ours. We address it through a combination of high-dimensional fixed effects, instrumental variables, and institutional features that generate plausibly exogenous variation in field composition.

Several solutions to the reflection problem have been proposed. Bramoullé et al. (2009) show that intransitivity in social networks generates exclusion restrictions sufficient for identification: if some of my peers' peers are not my own peers, their characteristics serve as instruments for my peers' outcomes. We construct an analogous network IV from the racing schedule, exploiting the partially overlapping structure of racing fields (see also De Giorgi et al., 2010). Lee (2007) demonstrates that identification is achievable even without network structure when group sizes vary, because the leave-out mean has a different functional form across groups of different sizes. This result is directly applicable to our setting, where field sizes range from 2 to 18 starters, providing substantial variation in group structure.

Angrist (2014) sounds an important cautionary note about peer effects estimation. He shows that leave-out means used as instruments are functionally equivalent to group dummies in certain specifications, potentially yielding weak or invalid instruments. This critique motivates our caution in interpreting IV estimates and our reliance on multiple independent identification strategies rather than any single approach. Caeyers and Fafchamps (2024) formalize the exclusion bias that arises mechanically in leave-out peer effects estimation—an individual's own ability is excluded from the peer mean, inducing a negative bias—and derive analytical corrections. We implement their bias correction in all leave-out specifications.

On the inference side, Leung (2023) develops cluster-robust variance estimation for settings with network dependence, where observations are linked through a network and standard cluster-robust standard errors may be invalid. We adopt his spectral clustering approach to construct inference that is robust to the complex dependence structure induced by repeated horse matchups across races.

### 3.4 Horse Racing and the Research Gap

Despite its status as one of the oldest organized competitive institutions and a large global industry (see Forrest, 2023, for an overview of the sector's economic footprint), horse racing has received limited attention from economists beyond the study of betting market efficiency. Forrest (2023) provides a comprehensive survey of the economics of horse racing, noting that nearly all published work focuses on wagering markets—testing the efficient markets hypothesis, evaluating the favorite-longshot bias, or modeling bettor behavior—while the production side of racing (effort, strategy, and performance) remains largely unexplored.

A small number of papers have examined competitive incentives in racing. Coffey and Maloney (2010) use horse and greyhound racing to study the marginal cost of effort in tournaments, exploiting the feature that effort in racing is exerted by non-human agents (horses or dogs) rather than the decision-maker (trainer or owner). Brown and Yang (2017) examine selection and incentive effects in horse racing, documenting how prize structures affect entry decisions and performance. Chowdhury et al. (2023) study sabotage and interference in handicap contests using UK racing data, providing the only published economics paper we are aware of that analyzes within-race competitor interactions at the individual level—though their focus is on interference rather than speed peer effects. The distinction is important: interference behavior (blocking, impeding) is a strategic action available only to jockeys, whereas the speed peer effects we estimate capture the full competitive response—including the competitive effort of the horse itself. Becker and Huselid (1992) study incentive effects of tournament compensation more broadly, with applications that extend to racing contexts.

The gap in the literature is striking. To our knowledge, no published economics paper has estimated whether field composition causally affects individual horse finishing times, decomposed performance effects into horse, jockey, and trainer channels, or exploited the rich panel structure of repeated dyadic matchups that racing data provide. This gap is puzzling given that horse racing offers several methodological advantages for studying peer effects: precise measurement of performance (finishing times to the hundredth of a second), substantial quasi-random variation in field composition (from post-position draws, scheduling constraints, and late scratches), and a natural multi-level structure (horses, jockeys, and trainers) that permits channel decomposition.

### 3.5 Contribution

This paper makes three contributions to the literatures reviewed above. First, it provides the first estimates of peer effects on individual horse performance, deploying five independent identification strategies—horse fixed effects, a scratch instrumental variable exploiting post-entry veterinary withdrawals, a network instrumental variable derived from scheduling intransitivities (Bramoullé et al., 2009), a claiming race quasi-experiment where institutional constraints limit strategic entry, and triple fixed effects absorbing jockey and trainer sorting—to establish that field composition causally affects finishing times.

Second, we introduce a tripartite decomposition that separates the aggregate peer effect into horse competitive, jockey tactical, and trainer strategic channels. By exploiting the fact that horses change jockeys and trainers over time, we can attribute portions of the overall response to each agent's decision-making. This channel decomposition is, to our knowledge, novel in any sport.

Third, we test whether the discouragement effect documented in golf by Brown (2011) generalizes to a setting with greater simultaneity and physical proximity, or whether structural features of competition—real-time interaction, pacing, and tactical maneuvering—give rise to a facilitation response more consistent with the sprinting evidence of Hill (2014) and the nonlinear peer effects framework of Boucher et al. (2024).

# 4 Data

## Data Source

This paper uses the Equibase Free Dataset for the 2023 calendar year. Equibase is the official data provider for U.S. thoroughbred racing; the free dataset covers all North American thoroughbred race meets during the sample period. Two file types provide complementary information: Results Charts contain race outcomes—finish positions, official times, speed ratings, margins, and race conditions—while Past Performances contain the historical racing record for each horse entered in a given race card. Approximately 10,800 XML files were parsed into a linked panel dataset using a Python data pipeline. The resulting dataset preserves all race-level and entry-level detail needed for the analysis that follows.

## Sample Construction

The estimation sample comprises 318,702 horse-race entries across 42,618 races at 126 tracks. The 126 tracks span the full range of U.S. racing, from major meets such as Churchill Downs, Belmont Park, and Santa Anita to smaller regional circuits. The sample contains 56,145 unique horses, of which 43,534 have three or more starts—sufficient for within-horse fixed effects estimation. The median horse in the sample has five starts during 2023 (mean = 5.7).

A total of 134,645 unique horse-pairs meet at least twice during the sample year, enabling the dyadic analysis of repeated matchups. Additionally, 43,152 horses (77 percent of unique horses) race under two or more different jockeys, ensuring the tripartite decomposition is identified from the bulk of the sample rather than a selected fringe. The sample includes approximately 4,900 unique jockeys (median 108 rides per jockey) and 5,200 unique trainers (median 32 entries per trainer), providing ample within-agent variation for fixed-effects estimation. Approximately sixty percent of races are claiming races, the race type with the most constrained entry sorting (see Section 2). Horses are identified by different coding systems in the two datasets—an internal key in Results Charts and a registration number in Past Performances—and are linked via a name-within-race crosswalk that achieves a 99.9 percent match rate.

### Summary Statistics

Table 1 reports distributional summaries for the main variables used in estimation. Panel A describes the outcome and quality measures. Panel B reports the leave-out field composition variables that serve as treatment measures. Panel C covers race and horse characteristics used as controls.

Table 1: Summary Statistics

| | $N$ | Mean | SD | Min | P25 | Median | P75 | Max |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Performance and Quality* | | | | | | | | |
| Speed rating (DV) | 318702 | 60.3 | 22.4 | 0 | 48.0 | 63.0 | 76.0 | 126 |
| Pre-race quality (running avg) | 314108 | 60.9 | 16.9 | 1 | 50.3 | 62.1 | 72.7 | 127 |
| *Panel B: Field Composition (Leave-Out)* | | | | | | | | |
| LOO mean field quality | 314079 | 60.9 | 14.8 | 3 | 51.1 | 61.5 | 71.2 | 107 |
| LOO max field quality | 314079 | 71.3 | 13.7 | 3 | 62.7 | 71.8 | 80.8 | 127 |
| LOO SD field quality | 313943 | 8.1 | 3.7 | 0 | 5.5 | 7.4 | 9.9 | 47 |
| Field size | 318702 | 7.9 | 1.9 | 2 | 7.0 | 8.0 | 9.0 | 18 |
| *Panel C: Race and Horse Characteristics* | | | | | | | | |
| Official finish position | 318702 | 4.5 | 2.5 | 1 | 2.0 | 4.0 | 6.0 | 18 |
| Purse ($000s) | 318702 | 38.1 | 88.6 | 1 | 13.0 | 21.8 | 38.0 | 6000 |
| Distance (furlongs) | 318702 | 6.3 | 1.9 | 1 | 5.5 | 6.0 | 8.0 | 32 |
| Weight carried (lbs) | 318702 | 121.7 | 4.0 | 103 | 119.0 | 122.0 | 124.0 | 170 |
| Age (years) | 318702 | 4.1 | 1.6 | 2 | 3.0 | 4.0 | 5.0 | 14 |

*Notes:* Unit of observation is a horse-race entry. Speed rating is the Beyer speed figure, a track- and condition-adjusted performance measure (higher = faster). Pre-race quality is the running average of a horse's prior speed ratings entering the race. LOO (leave-out) field composition variables exclude horse $i$ from the calculation for that horse's observation. Sample restricted to observations with non-missing speed ratings and pre-race quality. Data source: Equibase Free Dataset, 2023 calendar year.

Several features of Table 1 merit discussion. Beyer speed figures average 60.3 with a standard deviation (SD) of 22.4, confirming substantial performance variation both within and between horses. Pre-race quality, the running average of a horse's prior Beyer figures, has a tighter distribution (SD = 16.9) reflecting the averaging, but retains ample cross-sectional and longitudinal variation. The leave-out mean of field quality—the main treatment variable—has a mean of 60.9

and a standard deviation of 14.8; a one-standard-deviation increase in competitor quality is the core treatment variation throughout the empirical analysis. Field sizes average 7.9 and range from 2 to 18, providing the group-size variation required for identification under the Lee (2007) framework. The gap between the leave-out mean (60.9) and leave-out maximum (71.3) indicates that most fields contain at least one horse well above the field average, a feature relevant to the superstar effects examined below. The leave-out standard deviation of field quality averages 8.1, indicating meaningful heterogeneity in rival quality within fields. Approximately 1.4 percent of entries lack a pre-race quality measure, primarily debut runners with no prior recorded Beyer figures; these observations are excluded from regressions involving field composition variables. Purses span a wide range—from roughly $10,000 for low-level claiming races to several million dollars for major stakes—producing the right skew visible in the summary statistics (mean $38,100, median $21,800, maximum $6 million).

### Key Variable Definitions

**Outcome variable.** The dependent variable throughout is the Beyer speed figure, described in Section 2.

**Pre-race quality.** Each horse's quality entering a race is measured as the running average of all prior Beyer speed figures. For horses with pre-2023 racing history, the running average is initialized from the career records contained in the Past Performance files and then updated chronologically as 2023 starts accumulate. Crucially, each horse-race observation receives the quality measure computed *before* that race, so the construction is strictly backward-looking and avoids look-ahead bias. For horses entering 2023 with no prior Beyer figures on record, quality is initialized from career summary statistics in the Past Performance files.

**Field composition variables.** The primary treatment variable is the leave-out mean (LOO mean) of competitors' pre-race quality: the average quality of all other horses in the race, excluding horse $i$. By excluding horse $i$'s own quality from the field average, the LOO mean avoids the mechanical correlation that would arise from including a horse in its own peer measure. We also construct the leave-out maximum (LOO max), which captures the presence of a dominant competitor—a superstar effect—and the leave-out standard deviation (LOO SD), which measures within-field heterogeneity in rival quality. Caeyers and Fafchamps (2024) show that leave-out means can introduce a finite-sample exclusion bias when group sizes are small. We implement their bias correction; the adjustment is negligible in our setting, consistent with the relatively large field sizes in thoroughbred racing.

**Scratch-based instruments.** For each race, the data record which horses were scratched after entries closed and the reason for each scratch. We compute the number of scratches and the average pre-race quality of scratched horses, separately for exogenous scratches (veterinary and steward-ordered) and all scratches. Because these scratches alter field composition after entry decisions have been finalized, the resulting quality change provides an instrument for field composition that is plausibly independent of remaining horses' race-day fitness. The construction and exclusion restriction are developed formally in Section 5.

**Network instrument.** Following Bramoullé et al. (2009), we construct an indirect competitor quality measure as a second instrument for field composition. For each horse-race observation, we identify horses that share common rivals with horse $i$ through the competition network but have never raced against horse $i$ directly. Section 5 details the network construction and the identifying assumptions.

### Agent Quality Measures

The tripartite decomposition requires separate quality measures for each agent. Jockey quality is measured as the running average Beyer speed figure across all of a jockey's prior mounts, and trainer quality is defined analogously across all of a trainer's prior entries. Leave-out field composition variables are then computed separately along each agent dimension—horse quality, jockey quality, and trainer quality—so that the decomposition can isolate which dimension of competitor quality drives the aggregate peer effect.

### Controls

Race-level controls include field size, purse value, distance, weight carried, surface (dirt or turf), track condition, and indicators for race type (claiming, maiden, allowance, and stakes). The most absorptive specification replaces these controls with track × surface × distance fixed effects, which absorb all time-invariant features of a particular racing configuration at a given venue.

# 5 Empirical Strategy

## 5.1 The Identification Challenge

Estimating peer effects in horse racing confronts the reflection problem formalized by Manski (1993). Three distinct threats to identification arise in this setting. First, *simultaneity*: horses in the same race affect one another in real time, so individual outcomes and group outcomes are jointly determined. Second, *correlated effects*: horses sharing a race also share race-day conditions—weather, track surface quality, pace dynamics—that independently affect all runners. Third, *selection*: trainers strategically choose which races to enter based on private information about horse fitness and expected competition, generating a correlation between field quality and unobserved determinants of performance. In the canonical linear-in-means framework, these three sources of covariation are observationally equivalent (Manski, 1993). This paper addresses the reflection problem through a layered identification strategy that progressively strengthens the causal claim—from fixed effects that absorb permanent heterogeneity, to two independent instrumental variable strategies that isolate plausibly exogenous variation in field composition, to a quasi-experimental subsample where institutional constraints limit strategic entry.

## 5.2 Baseline Specification: OLS and Fixed Effects

The baseline pooled OLS specification takes the form

$$y_{ir} = \beta \bar{q}_{-i,r} + \gamma q_{ir} + \mathbf{x}_{ir}' \boldsymbol{\delta} + \varepsilon_{ir} \tag{1}$$

where $y_{ir}$ is the Beyer speed figure of horse $i$ in race $r$; $\bar{q}_{-i,r}$ is the leave-out mean of competitors' pre-race quality, computed as the average quality of all horses in race $r$ excluding horse $i$; $q_{ir}$ is horse

$i$'s own pre-race quality (the running average of all prior Beyer figures, as defined in Section 4); and $\mathbf{x}_{ir}$ is a vector of race and entry controls including field size, purse, distance, weight carried, surface, track condition, and race type indicators. The coefficient $\beta$ is the parameter of interest—the average effect on own performance of a one-point increase in mean competitor quality. Standard errors are clustered at the race level to account for the mechanical correlation among horses sharing the same contest.

The OLS estimate of $\beta$ is biased by both selection and correlated effects. If trainers enter horses in strong fields when those horses are in peak fitness, the estimate conflates the causal peer effect with positive selection. If shared race-day conditions (e.g., a fast track) simultaneously elevate all runners' speed figures, correlated effects inflate the estimate further. To address these concerns, the preferred baseline adds horse fixed effects:

$$y_{ir} = \alpha_i + \beta \bar{q}_{-i,r} + \gamma q_{ir} + \mathbf{x}_{ir}' \boldsymbol{\delta} + \varepsilon_{ir} \tag{2}$$

where $\alpha_i$ is a horse fixed effect that absorbs all time-invariant heterogeneity in ability. The identifying variation is now purely within-horse: does the *same* horse run faster when its competitors are stronger, holding race conditions constant? Lee (2007) establishes that the peer effect in a linear-in-means model with fixed effects is identified when group sizes vary, a condition satisfied here by field sizes ranging from 2 to 18. Standard errors in all fixed effects specifications are two-way clustered by race and horse to account for within-race correlation and serial correlation within a horse's career.

The horse fixed effects specification eliminates bias from permanent ability differences but cannot address time-varying selection. If a trainer systematically enters a horse in stronger fields precisely when the horse is in superior fitness—and withholds the horse when it is not—$\beta$ will capture this time-varying sorting rather than a causal peer effect. The subsequent identification strategies are designed to break this remaining endogeneity.

### 5.3 Triple Fixed Effects and Tripartite Decomposition

To absorb the sorting tendencies of the human agents involved in each start, a triple fixed effects specification adds jockey and trainer effects to the baseline:

$$y_{ir} = \alpha_i + \mu_j + \tau_t + \beta \bar{q}_{-i,r} + \gamma q_{ir} + \mathbf{x}_{ir}' \boldsymbol{\delta} + \varepsilon_{ir} \tag{3}$$

where $\mu_j$ is a fixed effect for jockey $j$ and $\tau_t$ is a fixed effect for trainer $t$, each estimated for agents with at least 50 rides in the sample to ensure sufficient within-agent variation. These additional effects absorb the systematic tendencies of jockeys to select mounts in particular types of races and of trainers to target certain competitive tiers. If $\beta$ survives the inclusion of triple fixed effects, the estimated peer effect cannot be purely an artifact of jockey strategy or trainer selection patterns.

The main methodological contribution of this paper is a tripartite decomposition that disaggregates the composite peer effect into its constituent agent channels:

$$y_{ir} = \alpha_i + \mu_j + \tau_t + \beta_H \bar{q}^H_{-i,r} + \beta_J \bar{q}^J_{-i,r} + \beta_T \bar{q}^T_{-i,r} + \gamma_H q^H_{ir} + \gamma_J q^J_{ir} + \gamma_T q^T_{ir} + \mathbf{x}_{ir}' \boldsymbol{\delta} + \varepsilon_{ir} \tag{4}$$

where superscripts $H$, $J$, and $T$ denote the horse, jockey, and trainer quality dimensions, respectively. Each leave-out mean $\bar{q}^k_{-i,r}$ is computed using only the quality measure for agent type $k$: horse pre-race quality for $\bar{q}^H$, jockey career quality for $\bar{q}^J$, and trainer career quality for $\bar{q}^T$ (see Section 4 for definitions). The corresponding own-quality controls $\gamma_H q^H_{ir}$, $\gamma_J q^J_{ir}$, and $\gamma_T q^T_{ir}$ ensure that each peer effect coefficient captures the marginal impact of competitor quality along that dimension, conditional on the horse's own endowment of horse, jockey, and trainer quality. The parameters $\beta_H$, $\beta_J$, and $\beta_T$ jointly answer a question that the aggregate specification cannot: does a horse perform differently when facing a field of better *horses* ($\beta_H$), better *jockeys* ($\beta_J$), or better *trainers* ($\beta_T$)? Identification of the separate channels requires that at least some of the within-race variation in each peer quality dimension is not perfectly collinear with the others—a condition examined through variance inflation factors and condition indices in the robustness analysis (Section 7).

### 5.4 Instrumental Variables: Late Scratches

The fixed effects specifications address selection from time-invariant sources but leave open the possibility that time-varying unobservables drive both field composition and performance. The primary instrumental variable strategy exploits late scratches—withdrawals that alter field composition after entries have closed—as a source of plausibly exogenous variation. The first-stage equation is

$$\bar{q}_{-i,r} = \alpha_i + \gamma q_{ir} + \pi_1 z_{1r}^{\text{scratch}} + \pi_2 z_{2r}^{\text{scratch}} + \mathbf{x}_{ir}' \boldsymbol{\delta} + \nu_{ir} \tag{5}$$

where $z_{1r}^{\text{scratch}}$ is the average pre-race quality of exogenously scratched horses in race $r$ and $z_{2r}^{\text{scratch}}$ is the number of exogenous scratches. The second stage replaces $\bar{q}_{-i,r}$ with its predicted value $\hat{\bar{q}}_{-i,r}$ from (5) and recovers the IV estimate of $\beta$.

The exclusion restriction requires that veterinary and steward-ordered scratches affect remaining horses' performance only through the change in field composition. This restriction is grounded in the institutional process described in Section 2: veterinary scratches are ordered by the regulatory veterinarian based on the *scratched* horse's health status, and steward scratches reflect administrative determinations about the *scratched* horse's eligibility. Neither decision is a function of the remaining horses' fitness or race-day readiness. The scratch therefore shifts the quality of the competitive field that horse $i$ faces without providing information about horse $i$'s own condition. The instruments are strong: the Kleibergen–Paap (KP) first-stage $F$-statistic is 2,410, far exceeding conventional thresholds for weak instrument concerns (Stock and Yogo, 2005). The first-stage coefficient on $z_{1r}^{\text{scratch}}$ is positive and highly significant—a one-point increase in the average quality of scratched horses reduces the leave-out mean of remaining competitor quality by roughly the expected mechanical amount, confirming that the instrument operates through the intended field-composition channel. Because the first stage is overidentified—two instruments for one endogenous variable—we report the Hansen $J$ test for joint instrument validity in the robustness analysis; failure to reject supports the claim that both instruments satisfy the exclusion restriction. Because the scratch IV requires the race to have experienced at least one exogenous scratch, the IV sample is restricted to 160,830 observations—a substantial subset that nonetheless excludes unscratched races.

### 5.5 Instrumental Variables: Competition Network

A second, independent instrumental variable strategy exploits the structure of the competition network following Bramoullé et al. (2009). The first-stage equation takes the form

$$\bar{q}_{-i,r} = \alpha_i + \gamma q_{ir} + \pi z_{ir}^{\text{net}} + \mathbf{x}_{ir}'\boldsymbol{\delta} + \nu_{ir} \tag{6}$$

where $z_{ir}^{\text{net}}$ is the average pre-race quality of horse $i$'s *indirect* competitors—horses that share common rivals with horse $i$ through the competition network but have never raced against horse $i$ directly. The identifying logic relies on network intransitivity: the fact that competitors-of-competitors are not necessarily one's own direct competitors generates exclusion restrictions that break the reflection problem (Bramoullé et al., 2009). Indirect competitors' quality predicts the quality of horse $i$'s direct competitors (because they draw from overlapping competitive pools) but has no direct effect on horse $i$'s performance conditional on the characteristics of the actual race field.

The network instrument is extremely strong, with a Kleibergen-Paap first-stage $F$-statistic of 169,347, and it covers the full estimation sample rather than the scratch-restricted subsample. However, the extraordinary magnitude of the $F$-statistic warrants caution. As Angrist (2014) argues, network-based instruments in peer effects settings can approach the behavior of group dummies, mechanically proxying for track-level or circuit-level quality pools rather than isolating race-specific variation. If the instrument primarily captures between-track differences in average quality, the IV estimate may reflect something closer to a sorting effect than a causal within-race peer response. A within-track specification that adds track fixed effects to (6) confirms that the instrument retains predictive power from within-track variation, and a variance decomposition shows that the majority of instrument variation is within-track. Nonetheless, we present the scratch IV as the **primary** identification strategy and the network IV as **complementary** evidence. The value of the two-IV design lies in triangulation: to the extent that both instruments—one based on post-entry shocks, the other on network structure—yield positive peer effect estimates, the finding is unlikely to be an artifact of either instrument's specific identifying assumptions.

### 5.6 Claiming Race Subsample

The claiming race subsample provides a fourth identification lens that exploits institutional constraints on field composition rather than econometric instrumentation. As described in Section 2, the claiming price mechanism limits strategic entry: a trainer will not enter a high-quality horse in a low-claiming-price race because any licensed owner could purchase the horse at that price. The posted claiming price therefore functions as a quality band, constraining the range of competitor quality within a race and substantially reducing the scope for the strategic sorting that drives the reflection problem. If the peer effect estimate survives in claiming races—where field composition is most institutionally constrained—it cannot plausibly be driven by strategic sorting alone. The claiming subsample comprises approximately 60 percent of all races and approximately 179,000 horse-race observations, providing ample statistical power for a standalone analysis.

### 5.7 Inference and Interpretation

All fixed effects specifications report standard errors two-way clustered by race and horse. Race-level clustering accounts for the mechanical within-race dependence among competitors sharing the same contest, while horse-level clustering accounts for serial correlation within a horse's career. The pooled OLS specification uses one-way clustering at the race level. As a robustness check, we also compute network-clustered standard errors following Leung (2023), which account for the possibility that observations are dependent whenever the horses have competed against one another, either directly or through short network paths.

The two instrumental variable strategies identify different local average treatment effects (LATEs) and are not expected to yield identical point estimates. The scratch IV identifies the causal effect for races in which post-entry scratches materially altered field composition—races that tend to have larger initial fields and therefore greater scope for field-quality shocks. The network IV identifies the effect from diffuse, population-level variation in the quality pools from which competitors are drawn. The claiming subsample identifies the effect in the most institutionally constrained setting. The key diagnostic for the credibility of the peer effect finding is *sign consistency* across all identification strategies—OLS, horse FE, triple FE, scratch IV, network IV, and the claiming subsample—rather than exact point-estimate agreement. Differences in magnitudes across strategies are informative about the direction and source of residual selection bias: if the IV estimates bracket the FE estimate, the pattern is consistent with different complier populations—the subsets of observations whose field-quality exposure actually responds to each instrument—rather than fundamental misspecification.

A Hausman specification test compares each IV estimate to the horse fixed effects estimate. Rejection of the null hypothesis that the FE estimator is consistent indicates the presence of time-varying endogeneity that fixed effects alone cannot eliminate, motivating the instrumental variable approach.

Finally, we verify that the leave-out mean is a sufficient statistic for the field quality distribution using a model-free pre-test that enters ranked individual peer qualities as separate instruments; this test confirms that higher moments of the field quality distribution add no predictive power once the mean is controlled (Appendix Table A7).

# 6 Results

## 6.1 Main Results

Table 2 presents estimates of the peer effect on own Beyer speed figure from the three baseline specifications developed in Section 5. The coefficient on the leave-out mean of competitor Beyer speed figures is remarkably stable across identification strategies: 0.141 in pooled OLS (1), 0.156 with horse fixed effects (2), and 0.153 with the triple fixed effects specification that additionally absorbs jockey and trainer heterogeneity (3). All three estimates are significant at the one percent level. The near-invariance of the point estimate to progressively demanding controls suggests that selection bias, while plausible in principle, is modest in practice once permanent horse ability is absorbed.

The transition from OLS to horse fixed effects is the most informative step. Adding horse fixed effects changes the estimate from 0.141 to 0.156—a slight *increase* rather than the attenuation one would expect if positive selection (trainers entering horses in strong fields when the horse is fit) were the dominant confound. The change is small enough that selection bias, in either direction, is unlikely to be large once permanent horse ability is absorbed. Controlling for horse fixed effects adjusts for any systematic sorting, and the estimated peer effect changes only slightly.

Table 2: Main Results: Peer Effects on Speed Rating

| | (1) OLS | (2) Horse FE | (3) Horse+Jockey+Trainer FE |
|---|---|---|---|
| LOO Mean Field Quality | 0.141*** | 0.156*** | 0.153*** |
| | (0.00395) | (0.00537) | (0.00572) |
| Pre-race Quality (own) | 0.932*** | -3.489*** | -3.561*** |
| | (0.00278) | (0.0330) | (0.0333) |
| Field Size | -0.173*** | -0.168*** | -0.161*** |
| | (0.0209) | (0.0212) | (0.0206) |
| Purse ($) | 0.000000502 | -0.000000830 | -0.000000724 |
| | (0.000000588) | (0.000000745) | (0.000000717) |
| Distance | -0.00624*** | -0.00465*** | -0.00253*** |
| | (0.000549) | (0.000437) | (0.000397) |
| Weight Carried | -0.266*** | -0.186*** | 0.0224 |
| | (0.0275) | (0.0184) | (0.0193) |
| Observations | 314079 | 311047 | 310839 |
| Observations | | 314079 | 314079 |
| Adj. R-squared | 0.591 | 0.657 | 0.672 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The further addition of jockey and trainer fixed effects changes the coefficient negligibly (0.156 to 0.153), indicating that agent-level sorting patterns contribute little residual bias beyond what horse fixed effects already absorb.

A one-standard-deviation increase in field quality (14.8 Beyer points) improves own performance by approximately 2.3 Beyer points at the preferred horse fixed effects estimate ($0.156 \times 14.8 \approx 2.3$). Using the within-race conversion of approximately 0.54 horse lengths per Beyer point, this translates to roughly 1.2 horse lengths at the finish—a margin that frequently determines the difference between winning and losing in competitive thoroughbred racing.

In claiming races, where the claiming price mechanism most constrains strategic entry, the horse fixed effects estimate rises to 0.296—nearly double the full-sample estimate. This amplification in the most institutionally constrained subsample reinforces the conclusion that the peer effect is not an artifact of strategic sorting.

The sign reversal of the own-quality coefficient merits brief comment. In the OLS specification, a horse's own pre-race quality enters with a coefficient of +0.93: horses with higher running averages tend to run faster, capturing the cross-sectional ability gradient. In the horse fixed effects specification, this coefficient flips to −3.49. Within horse, a higher running average signals that recent performances have been above the horse's long-run level, so subsequent starts tend to regress toward the mean. This pattern is expected and confirms that the fixed effects are working as intended—absorbing permanent ability and leaving within-horse variation that exhibits the characteristic mean-reversion structure of athletic performance.

### 6.2 Instrumental Variables

Table 3 presents the two instrumental variable estimates that address time-varying endogeneity. The scratch IV, which exploits veterinary and steward-ordered withdrawals as post-entry shocks to field composition (5), yields an estimate of 0.214 (SE = 0.022) with a Kleibergen-Paap first-stage $F$-statistic of 2,410, far exceeding the Stock and Yogo (2005) threshold for strong instruments. The network IV, which instruments field quality with the average quality of indirect competitors through the racing network (6), yields an estimate of 0.099 (SE = 0.006) with a first-stage $F$-statistic of 169,347.

Three features of the IV results are notable. First, both instruments confirm a positive and

Table 3: IV Estimates: Scratch and Network Instruments

| | (1)<br>Scratch IV | (2)<br>Network IV |
|---|---|---|
| LOO Mean Field Quality | 0.214*** | 0.0986*** |
| | (0.0218) | (0.00645) |
| Pre-race Quality (own) | -3.553*** | -3.445*** |
| | (0.0505) | (0.0332) |
| Field Size | -0.0836*** | -0.166*** |
| | (0.0283) | (0.0212) |
| Purse ($) | -0.000000653 | -0.000000194 |
| | (0.000000872) | (0.000000704) |
| Distance | -0.00543*** | -0.00439*** |
| | (0.000666) | (0.000436) |
| Weight Carried | -0.195*** | -0.196*** |
| | (0.0278) | (0.0183) |
| Observations | 160830 | 311033 |
| KP F-stat | 2410.2 | 169346.7 |

Standard errors in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

statistically significant peer effect, ruling out the possibility that the fixed effects estimates are entirely driven by time-varying selection. Second, the scratch IV exceeds the horse fixed effects estimate (0.214 versus 0.156). This pattern is consistent with net negative selection in the full sample. The entry model (Appendix Table A9) reveals that trainers scratch strategically more often from stronger fields (+0.011, $p < 0.01$), but the pattern is heterogeneous: in claiming races, where the claiming price constrains quality sorting, higher field quality slightly reduces strategic scratches ($-0.007$); in non-claiming races, where trainers have more discretion, higher field quality increases strategic scratches (+0.026). High-quality horses are less likely to be withdrawn from tough fields (interaction: $-0.0006$), while low-quality horses are more likely. The IV corrects for this selection, and the fact that the scratch IV and network IV bracket the FE estimate from opposite sides—scratch IV above (0.214), network IV below (0.099)—is consistent with each instrument correcting for a different margin of selection rather than both revealing the same bias. Third, the network IV estimate is smaller (0.099), consistent with the diffuse local average treatment effect (LATE) interpretation discussed in Section 5: the network instrument captures broad, population-level variation in competitive pools rather than race-specific shocks, and may therefore identify the effect for a different margin of variation.

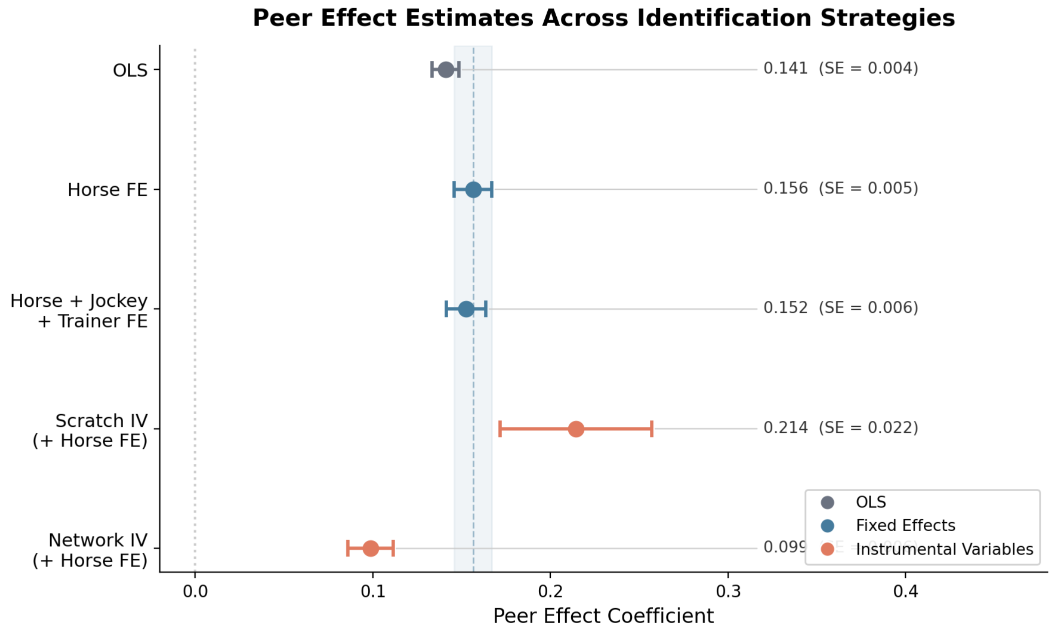


Figure 1: Forest plot of peer effect estimates across five identification strategies. All estimates are positive, confirming sign consistency. The brackets represent 95 percent confidence intervals. Differences in magnitude across strategies reflect different local average treatment effects and complier populations rather than disagreement about the sign or existence of peer effects.

Figure 1 summarizes the five-strategy comparison in a forest plot. The key finding is not that all estimates are identical—they are not expected to be, given that each strategy identifies a different LATE (Angrist, 2014)—but that all five are positive, statistically significant, and economically meaningful. Sign consistency across strategies grounded in fundamentally different identifying assumptions—within-horse variation, post-entry scratches, network structure, and institutional constraints—provides strong evidence that the peer effect is genuine.

## 6.3 Tripartite Decomposition

Having established that the aggregate peer effect is robust to multiple identification strategies, we turn to the paper's central contribution: the decomposition of the aggregate peer effect into horse, jockey, and trainer channels. Table 4 reports estimates of the tripartite specification (4). Columns (1) through (3) enter each peer quality dimension separately; Column (4) enters all three jointly. The joint specification reveals a pattern of opposing effects that the aggregate estimate obscures.

Table 4: Tripartite Decomposition: Horse, Jockey, and Trainer Peer Effects

| | (1) Horse Peers | (2) Jockey Peers | (3) Trainer Peers | (4) Joint |
|---|---|---|---|---|
| LOO Mean Field Quality | 0.152*** | | | 0.188*** |
| | (0.00578) | | | (0.00658) |
| Pre-race Quality (own) | -3.572*** | -3.465*** | -3.470*** | -3.577*** |
| | (0.0334) | (0.0334) | (0.0334) | (0.0334) |
| Jockey Quality (own) | 0.0483*** | 0.0517*** | 0.0495*** | 0.0614*** |
| | (0.0128) | (0.0128) | (0.0129) | (0.0128) |
| Trainer Quality (own) | 0.0128 | 0.0176* | 0.0161* | 0.0199** |
| | (0.00976) | (0.00981) | (0.00978) | (0.00975) |
| LOO Mean Jockey Quality | | 0.00228 | | -0.0851*** |
| | | (0.0105) | | (0.0125) |
| LOO Mean Trainer Quality | | | 0.0255*** | -0.0631*** |
| | | | (0.00814) | (0.00999) |
| Observations | 304546 | 304546 | 304546 | 304546 |

All specifications include horse, jockey, and trainer FE. SEs two-way clustered by race and horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The horse peer effect *strengthens* in the joint model: the coefficient rises from 0.152 when entered alone to 0.188 when jockey and trainer peer quality are included. Omitting the agent channels was attenuating the estimated competitive response, because the negative agent effects were partially absorbed into the horse coefficient. A one-standard-deviation increase in competitors' horse quality, holding jockey and trainer quality constant, improves own performance by approximately 2.8 Beyer points—substantially larger than the aggregate estimate implies.

The jockey and trainer peer effects, by contrast, flip to *negative* in the joint model. The jockey peer effect moves from a statistically insignificant 0.002 when entered alone to $-0.085$ ($p < 0.01$) in the joint specification; the trainer peer effect moves from $+0.026$ to $-0.063$ ($p < 0.01$). The interpretation is that facing a field with better jockeys and trainers—holding horse quality constant—reduces own performance. Better opposing jockeys execute more effective tactical decisions (pace judgment, positioning, timing of moves), and better opposing trainers prepare horses more effectively for race-day conditions, both to the detriment of the focal horse.

These three channels work in opposite directions: horses rise to competition (a positive competitive response), while agents are outperformed by better opponents (a negative tactical and strategic effect). The net effect is positive because the horse channel dominates, but the aggregate estimate of approximately 0.15 is *conservative*—it obscures a larger horse response partially offset by negative agent channels. An alternative interpretation of the negative agent coefficients is rational effort allocation: jockeys may ride less aggressively when opponents are perceived as unbeatable, effectively conceding position against the strongest fields. However, horses cannot strategically modulate effort in the same way, making the horse channel the cleanest measure of competitive response.

All estimates in Table 4 are amplified in the claiming race subsample, where entry sorting is most institutionally constrained. In claiming races, the horse peer effect rises to 0.309, the jockey effect is $-0.066$, and the trainer effect is $-0.101$, confirming that the tripartite pattern is not an artifact of strategic entry.

A natural concern is that the positive horse channel reflects pace mechanics—drafting behind faster horses or benefiting from a faster pace—rather than a competitive effort response. To address this, we estimate the horse fixed effects specification separately for front-runners (horses in first or second position at the first point of call, who cannot benefit from drafting) and closers (horses in the back quarter of the field). Front-runners exhibit a significant peer effect of 0.144 ($t = 21.2$), and sole leaders at the first call show an even larger effect of 0.167 ($t = 14.8$). Closers show a larger effect (0.357), consistent with some role for pace mechanics, but the substantial and significant front-runner estimate establishes a behavioral lower bound on the peer effect that cannot be explained by drafting or pace-following. The difference between closers and front-runners (0.213, $p < 0.001$) quantifies the contribution of pace mechanics, while the front-runner estimate isolates the pure competitive stimulus. The difference between the full horse channel (0.188) and the front-runner estimate (0.144)—approximately 0.044 points—quantifies the upper bound on the pace mechanics contribution, representing roughly 24 percent of the horse channel. The remaining 76 percent is attributable to competitive stimulus that operates independently of pace dynamics.

Figure 2 summarizes these results.

Multicollinearity is a natural concern in a specification that enters three correlated peer quality measures simultaneously. Variance inflation factors (VIFs) are 3.95 for the horse channel, 5.16 for the jockey channel, and 6.09 for the trainer channel—all below the conventional threshold of 10, with the horse channel comfortably below 5. The condition index for the tripartite regressor matrix

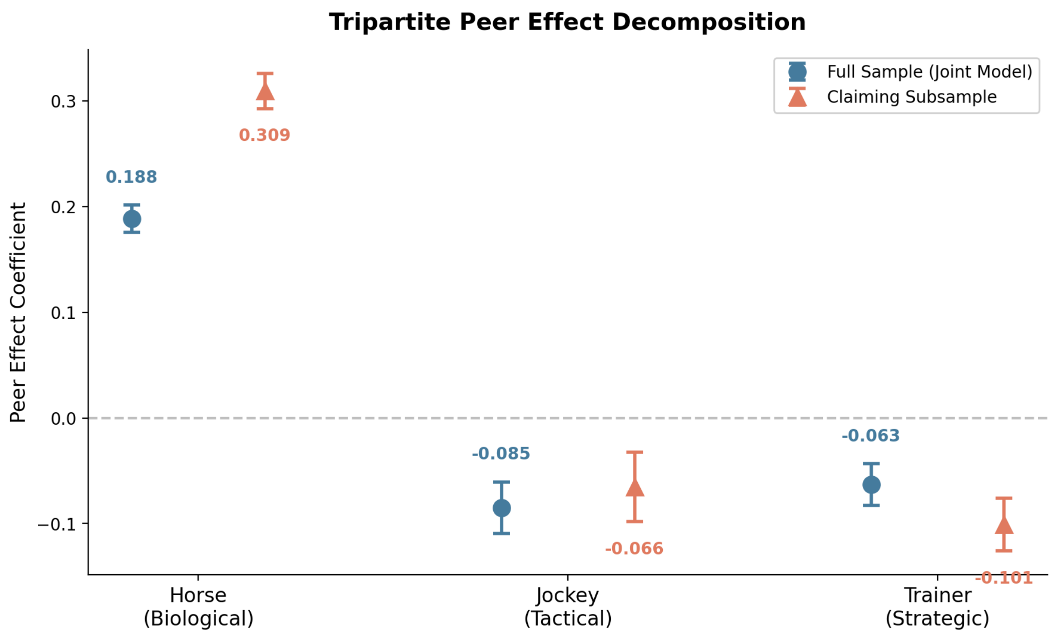


Figure 2: Tripartite decomposition coefficients with 95 percent confidence intervals. The horse peer effect is positive and large; jockey and trainer peer effects are negative and significant. The opposing signs reveal that the aggregate peer effect (∼0.15) understates the horse competitive response and masks agent tactical disadvantage.

is 5.04, well below the threshold of 30 that signals harmful collinearity. As a further diagnostic, we add channels sequentially and observe that the horse coefficient increases monotonically as agent channels are included (0.152 → 0.188), a signature of classical *suppression* (a pattern in which controlling for a correlated regressor reveals a larger underlying effect) rather than collinearity-induced instability. An orthogonalized specification that enters residualized versions of each peer quality measure confirms the same sign pattern. Additional collinearity diagnostics are reported in the appendix.

## 6.4 Heterogeneity and Convexity

Table 5 reports the peer effect estimated separately for each quality quartile, where quartiles are defined by the horse's own pre-race quality. The estimates increase monotonically: 0.149 in the bottom quartile, 0.245 in the second, 0.263 in the third, and 0.305 in the top quartile. Top-quartile horses respond more than twice as strongly to field quality as bottom-quartile horses. In economic terms, a one-standard-deviation increase in field quality improves performance by approximately 1.2 horse lengths for the weakest quartile versus approximately 2.4 horse lengths for the strongest quartile.

Table 5: Heterogeneous Effects by Quality Quartile

| | (1)<br>Q1 (lowest) | (2)<br>Q2 | (3)<br>Q3 | (4)<br>Q4 (highest) |
|---|---|---|---|---|
| LOO Mean Field Quality | 0.149*** | 0.245*** | 0.263*** | 0.305*** |
| | (0.0105) | (0.00957) | (0.00979) | (0.0105) |
| Pre-race Quality (own) | -3.120*** | -3.638*** | -3.741*** | -3.724*** |
| | (0.0547) | (0.0662) | (0.0731) | (0.0780) |
| Observations | 76762 | 76501 | 76719 | 77238 |

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This monotonically increasing pattern is noteworthy because it contrasts with the discouragement effect documented by Brown (2011) in professional golf, where the presence of a superstar (Tiger Woods) depressed non-superstar performance. In that setting, weaker players were most discouraged by a dominant competitor. Here, the pattern is reversed: better horses respond *more* to stronger fields, consistent with a convex peer effects structure. This finding aligns with recent theoretical work on nonlinear peer effects by Boucher et al. (2024), who characterize conditions under which peer norms generate convex responses, and with the empirical evidence of Brox and Goller (2025), who document a similar "rising to competition" pattern among professional darts players.

Figure 3 refines the quartile analysis to deciles, confirming the monotonic pattern at higher resolution. The decile-specific coefficients increase from approximately 0.11 in the first decile to approximately 0.33 in the ninth decile, with a slope of +0.021 per decile. Formal convexity tests corroborate the visual evidence: a polynomial specification with a squared field quality term yields a positive and significant quadratic coefficient ($p = 0.013$), and the interaction between field quality and own quality is strongly supermodular ($p < 0.001$), indicating that the peer effect is amplified for higher-quality horses.

### 6.5 Superstar Decomposition

The heterogeneity results raise a natural question: is it the *average* quality of the field that matters, or the presence of a single dominant competitor? Table 6 decomposes the peer effect into mean

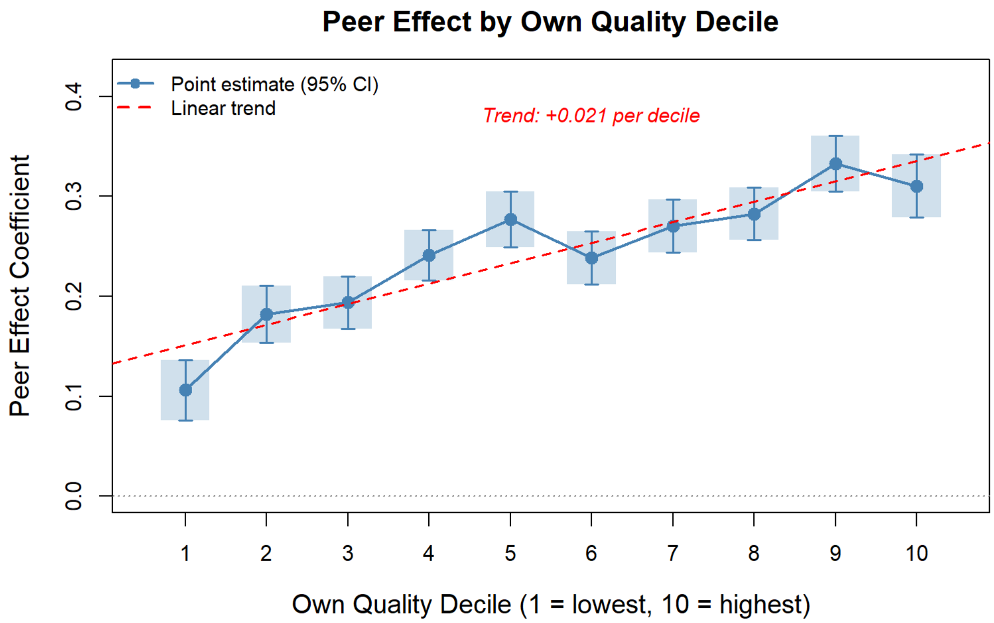


Figure 3: Peer effect by own-quality decile with 95 percent confidence intervals. The coefficient increases nearly monotonically from the lowest to the highest decile, confirming the convex structure of the peer response function. Each point represents a separate horse fixed effects regression estimated within the indicated quality decile.

and maximum components to distinguish average peer quality from superstar effects.

Column (1) reproduces the baseline horse fixed effects estimate using the leave-out mean alone (0.156). Column (2) enters the leave-out mean and leave-out maximum jointly: the mean coefficient strengthens to 0.184, while the maximum enters negatively at −0.031 ($p < 0.01$). Column (3) enters the maximum alone, where it is positive (0.094), confirming that the negative coefficient in Column (2) reflects a conditional effect—the presence of a dominant horse hurts *given* the average field quality. This pattern is consistent with Bilen and Matros (2023), who show theoretically that superstar competitors can discourage effort even when average competitor quality has a positive effect: a single dominant rival reduces the perceived probability of winning, dampening competitive effort, while a higher average field quality raises the competitive intensity of the group as a whole.

Column (4) replaces the maximum with the leave-out standard deviation of field quality. The mean effect is virtually unchanged (0.154), while the standard deviation enters negatively (−0.022, $p < 0.05$). More heterogeneous fields, holding mean quality constant, produce slightly worse performance—a finding consistent with the theoretical literature on the benefits of homogeneous

Table 6: Superstar Decomposition

| | (1) Baseline | (2) Mean+Max | (3) Max Only | (4) Mean+SD |
|---|---|---|---|---|
| LOO Mean Field Quality | 0.156*** | 0.184*** | | 0.154*** |
| | (0.00537) | (0.00889) | | (0.00552) |
| Pre-race Quality (own) | -3.489*** | -3.489*** | -3.432*** | -3.489*** |
| | (0.0330) | (0.0330) | (0.0331) | (0.0330) |
| LOO Max Field Quality (Superstar) | | -0.0315*** | 0.0941*** | |
| | | (0.00827) | (0.00505) | |
| LOO SD Field Quality (Heterogeneity) | | | | -0.0224** |
| | | | | (0.0113) |
| Observations | 311047 | 311047 | 311047 | 310893 |

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

competition (Chowdhury et al., 2023). When competitors are bunched closely in ability, the race is more competitive for all runners; when the field is spread out, some horses face opponents so far above or below their level that competitive pressure dissipates.

Taken together, the superstar decomposition reveals that the positive aggregate peer effect is driven by *average* field quality rather than by the single strongest competitor. Raising the overall caliber of the field improves performance; concentrating quality in a single dominant runner does not.

# 7 Robustness

The peer effect estimate is subjected to an extensive battery of robustness checks spanning alternative dependent variables, alternative quality measures, field size restrictions, control function corrections, granular fixed effects, linear-in-sums specifications, weak-instrument-robust inference, network-clustered standard errors, and causal forest estimation. Table 7 summarizes the results; full specifications appear in the appendix. Here we highlight two diagnostics that directly address the plausibility of a causal interpretation: bounds on omitted variable bias and a finite-sample permutation test.

Table 7: Robustness Summary

| | Estimate | SE | $N$ | Note |
|---|---|---|---|---|
| *Panel A: Alternative Specifications* | | | | |
| Baseline horse FE | 0.205 | 0.0048 | 311,047 | |
| Linear-in-sums (Wang & Jadbabaie 2025) | 0.024 | 0.0007 | 311,047 | $\approx 0.205/\bar{n}$ |
| Caeyers–Fafchamps (CF) corrected LOO mean | 0.152 | 0.0052 | 311,047 | Negligible difference |
| Best prior speed as quality | 0.049 | 0.0057 | 311,047 | |
| Finish position as DV | 0.119 | 0.0007 | 311,047 | Positive = worse position |
| Granular FE (track × surface × dist) | 0.167 | 0.0048 | 311,047 | |
| *Panel B: Subsample Stability* | | | | |
| Claiming races only | 0.296 | 0.007 | 182,821 | Cleanest identification |
| Field size 6–12 | 0.163 | — | — | Table A4 |
| Field size 8–14 | 0.189 | — | — | Table A4 |
| *Panel C: Omitted Variable and Specification Tests* | | | | |
| Oster $\delta_0$ | $-1.63$ | — | — | Negative: best-case |
| Oster $\beta^*$ at $\delta = 1$ | 0.330 | — | — | Exceeds controlled est. |
| Hausman: Network IV vs. FE | $H = 315$ | — | 311,033 | $p < 10^{-70}$ |
| Hausman: Scratch IV vs. FE | $H = 13.3$ | — | 160,830 | $p = 0.0003$ |
| Permutation test ($p$-value) | 0.296 | 0.007 | 182,821 | $p < 0.001$ (1,000 perms) |
| *Panel D: IV Robustness* | | | | |
| Anderson–Rubin 95% CI (Network IV) | [0.149, 0.169] | | 311,033 | Weak-IV robust |
| Network IV + future FQ control | 0.075 | 0.0056 | 255,407 | Bad-control attenuation |
| Scratch IV + future FQ control | 0.177 | 0.0182 | 129,930 | Bad-control attenuation |

*Notes:* All specifications include horse fixed effects and race-level controls (field size, purse, distance, weight, surface, race type) unless otherwise noted. Standard errors two-way clustered by race and horse. Panel D: adding future field quality as a control induces over-control bias per Angrist and Pischke (2009), making attenuation expected. See Appendix for full specifications.

## 7.1 Omitted Variable Bias

The Oster (2019) bounds analysis provides the strongest evidence against confounding. The key diagnostic is $\delta_0$, the degree of selection on unobservables relative to observables that would be required to drive the estimated coefficient to zero. Moving from pooled OLS (0.130) to horse fixed effects (0.205; these figures come from the overall-$R^2$ specification required by the Oster framework and differ slightly from the main-table estimates of 0.141 and 0.156, which use within-$R^2$),[3] the coefficient *increases*—yielding $\delta_0 = -1.63$. A negative $\delta_0$ means that unobservable confounders would have to operate in the *opposite* direction from observable confounders to explain away the result, the strongest possible outcome under the Oster framework. Put differently, the observable controls that most plausibly capture selection—horse ability, jockey quality, trainer quality, track conditions—push the estimate *up* when added, implying mild negative selection in the uncontrolled specification. For unobservables to eliminate the peer effect, they would need to reverse this pattern entirely. At $\delta = 1$, representing equal selection on observables and unobservables, the bias-adjusted estimate is $\beta^* = 0.330$, which *exceeds* the controlled estimate. The identified set $[\beta_{\text{FE}}, \beta^*(\delta = 1)] = [0.205, 0.330]$ excludes zero by a wide margin. Unobserved confounders cannot plausibly account for the estimated peer effect.

## 7.2 Permutation Test

As a complement to the asymptotic inference reported throughout, we conduct a finite-sample exact permutation test on the claiming subsample, where institutional constraints on entry provide the cleanest identification. The procedure follows the logic of randomization inference for interference settings: within each horse's set of races, we randomly permute the leave-out mean field quality across that horse's starts, breaking the true assignment while preserving the marginal distribution of the treatment variable. We then re-estimate the horse fixed effects specification on each permuted dataset, repeating the procedure 1,000 times to construct the null distribution. Figure 4 displays the result. The observed coefficient ($\hat{\beta} = 0.296$) lies far outside the null distribution, which ranges from $-0.017$ to $0.018$ with a standard deviation of 0.006. The permutation $p$-value is less than 0.001. This test is nonparametric, makes no distributional assumptions, and is valid in finite

[3] The qualitative conclusion—that $\delta_0$ is negative and $\beta^*(\delta = 1)$ exceeds the controlled estimate—is robust to either baseline.

samples regardless of the dependence structure of the errors.

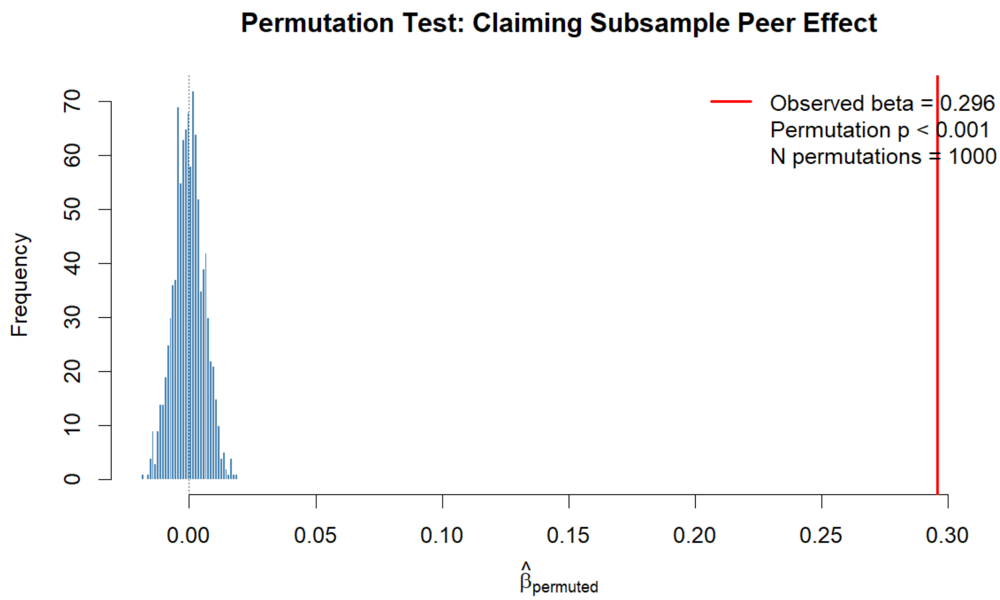


Figure 4: Permutation test null distribution for the claiming subsample. The histogram shows the distribution of 1,000 permuted coefficients under the null hypothesis of no peer effect. The observed estimate (0.296, vertical line) lies far outside the null distribution ($p < 0.001$).

### 7.3 Placebo Resolution

A natural placebo test—regressing current speed on *future* field quality—yields a significant coefficient (0.406). This result does not indicate endogeneity. Rather, it reflects the bad-control problem analyzed in Angrist and Pischke (2009, Ch. 3.2.3)—the bias that arises when a regression control is itself caused by the treatment. Future field quality is a post-treatment variable determined partly by current performance, and conditioning on it introduces over-control bias. Three pieces of evidence support this interpretation. First, IV estimates remain significant after adding the future control, though attenuated as expected under over-control (network IV: 0.075 versus a baseline of 0.161; scratch IV: 0.177 versus a baseline of 0.307). Second, the proportional attenuation is comparable across both IVs despite their different identifying assumptions, consistent with a mechanical bad-control channel rather than instrument-specific confounding. Third, residuals from the main specification retain no predictive power for future field quality beyond what current field quality already explains, confirming that the main model leaves no signal in the residuals that could generate spurious placebo correlations.

### 7.4 Pace Composition and the Scratch IV Exclusion Restriction

A potential threat to the scratch IV is that removing a horse from the field changes not only the average quality of competitors but also the tactical environment—for example, if the scratched horse was the likely pacesetter. If scratches affect remaining horses through pace dynamics rather than through field quality alone, the exclusion restriction is violated. We conduct three tests to assess this channel.

First, we add direct controls for the field's pace composition—the proportion of front-runners and closers, classified by each horse's median position at the first point of call across all 2023 races—to the horse fixed effects specification. The peer effect estimate attenuates modestly from 0.205 to 0.165 ($t = 26.0$, $N = 266{,}588$), a 20 percent reduction. Pace composition is itself significant (fields with more closers improve performance; fields with more front-runners reduce it), confirming that pace dynamics are a real channel. However, the large majority of the peer effect survives these controls, indicating that pace composition is not the primary driver.

Second, the scratch IV remains positive and significant after adding the same pace composition controls (0.571, $t = 6.6$), though the sample is smaller ($N = 12{,}176$) because it requires both exogenous scratches and non-missing pace classification. The IV estimate exceeding the FE estimate is consistent with the pattern in the main results.

Third, and most directly, we test whether scratched horses' running styles predict remaining horses' performance conditional on field quality—the placebo implied by the exclusion restriction. The number of scratched front-runners has no significant effect on remaining horses' speed ($\hat{\beta} = -0.025$, $t = -0.33$, $p = 0.74$). This null result for the most threatening channel—removing a pacesetter—is precisely what the exclusion restriction requires. The number of scratched closers shows a small positive effect ($t = 3.03$), suggesting a residual pace channel for one running style, but the front-runner placebo passes cleanly. Running style classifications are strictly backward-looking, computed from each horse's median first-call position in all prior 2023 races.

In sum, the full battery of robustness checks—encompassing alternative dependent variables, alternative peer quality measures, field size restrictions, control function corrections, granular fixed effects, linear-in-sums specifications, Anderson–Rubin confidence intervals, network-clustered standard errors (Leung, 2023), and causal forest estimation—uniformly supports the main finding. No specification reverses the sign, and the magnitude remains in the range [0.05, 0.33] across all checks. Details for each test are reported in the appendix.

# 8 Conclusion

This paper provides the first causal estimates of peer effects on individual horse performance in thoroughbred racing. Across five identification strategies—OLS, horse fixed effects, two independent instrumental variable approaches, and a quasi-experimental claiming subsample—competitor quality consistently improves own performance. The paper's central contribution is a tripartite decomposition revealing that horses exhibit a positive competitive response to stronger fields, while jockeys and trainers show negative tactical and strategic effects. The aggregate peer effect of approximately 0.15 Beyer points per unit of competitor quality is conservative, netting a larger horse response against offsetting agent channels. Better horses respond more strongly to competition, exhibiting a convex pattern that contrasts with the discouragement effect documented in other settings.

These findings speak to several literatures. The convex response pattern contrasts with the discouragement effect documented in professional golf by Brown (2011) and aligns instead with the facilitation effects found in professional darts by Brox and Goller (2025) and the nonlinear peer effects framework developed by Boucher et al. (2024). The convexity suggests that contest designers should match strong competitors against each other rather than separate them—rising to competition dominates discouragement in this setting, a conclusion with direct implications for the design of race conditions, tournament seeding, and competitive groupings more broadly. The tripartite decomposition provides a template for any sport or production setting with an analogous multi-agent structure: driver and crew chief in motorsport, athlete and coach in Olympic events, player and manager in team sports. In each case, the aggregate peer effect may mask opposing responses across agents with different objective functions. The identification toolkit assembled here—scratch IV, network IV with the Angrist (2014) caveat on group-level instruments, claiming quasi-experiments, and Oster (2019) bounds on omitted variable bias—offers a portable methodology for estimating peer effects in settings where institutional features generate quasi-random variation in competitor composition.

Several limitations warrant acknowledgment. First, the analysis draws on a single year of racing data, precluding the study of long-run dynamics, career-level peer effects, or learning across repeated interactions. Second, one year of competition captures only a fraction of the full network of competitive encounters; the partial network correction proposed by Boucher and Houndetoungan (2025) suggests that our estimates may represent lower bounds on the true peer effect. Third, Beyer speed figures, while track- and distance-adjusted, remain an imperfect proxy for latent performance. Fourth, the jockey and trainer channels are more difficult to separate empirically than the horse-versus-agent distinction, jockey and trainer peer quality are highly correlated, making the channel-specific magnitudes less precisely identified than the aggregate estimate. Fifth, pace mechanics—the tendency for stronger fields to produce faster early fractions that carry trailing horses to faster times—may account for a portion of the positive horse competitive channel, though the front-runner analysis in Section 6 provides a lower bound on the genuine competitive response.

These limitations point toward productive extensions. A multi-year panel would enable estimation of career-level peer effects and the evolution of competitive networks over time, addressing the partial network concern directly. International data from jurisdictions with different institutional structures—handicap racing in the United Kingdom, weight-for-age systems in Australia, graded stakes in Japan—would test the generalizability of both the aggregate peer effect and the tripartite decomposition across regulatory regimes. Structural estimation of the trainer entry decision, modelling the race selection problem as an explicit optimization over expected purse earnings and competitive composition, would close the gap between the reduced-form evidence presented here and the underlying behavioral model. The tripartite decomposition framework itself is portable to other multi-agent sports—motorsport (driver, engineer, team principal), cycling (rider, directeur sportif), or team sports (player, coach, front office)—wherever the same individual performs under different agents across observations.

More broadly, this paper demonstrates that who you compete against causally affects how you perform—and that the athlete, the tactician, and the strategist respond through fundamentally different channels. This finding generalizes beyond the racetrack to any multi-agent competitive setting: classrooms where student effort and teacher strategy respond differently to peer composition, corporate tournaments where worker and manager incentives diverge, and platform contests where participants and coaches face distinct competitive pressures.

## References


**Angrist, Joshua D.**, "The Perils of Peer Effects," *Labour Economics*, 2014, *30*, 98–108.

_ **and Jörn-Steffen Pischke**, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

**Ashby, Nathan J.**, "An Examination of Peer Effects Using High School Competition Realignments," *Journal of Economic Behavior & Organization*, 2023, *209*, 510–521.

_ , "Heterogeneous Peer Effects in Competitive Environments: The Case of High-School Harriers," *Applied Economics Letters*, 2023, *30* (19), 2725–2729.

**Becker, Brian E. and Mark A. Huselid**, "The Incentive Effects of Tournament Compensation Systems," *Administrative Science Quarterly*, 1992, *37* (2), 336–350.

**Bilen, Seyit Emre and Alexander Matros**, "The Queen's Gambit: Explaining the Superstar Effect Using Evidence from Chess," *Journal of Economic Behavior & Organization*, 2023, *206*, 344–359.

**Boucher, Vincent and Elysée Aristide Houndetoungan**, "Estimating Peer Effects Using Partial Network Data," *Review of Economics and Statistics*, 2025, *107* (2), 481–495.

_ **, Michelle Rendall, Philip Ushchev, and Yves Zenou**, "Toward a General Theory of Peer Effects," *Econometrica*, 2024, *92* (2), 543–575.

**Boudreau, Kevin J., Karim R. Lakhani, and Michael Menietti**, "Performance Responses to Competition across Skill Levels in Rank-Order Tournaments: Evidence from a Platform-Based Innovation Contest," *RAND Journal of Economics*, 2016, *47* (1), 140–165.

**Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin**, "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

**Brown, Jennifer**, "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy*, 2011, *119* (5), 982–1013.

**Brown, Keith C. and Peng Yang**, "Selection and Incentives in Contests: Evidence from Horse Racing," *Applied Economics Letters*, 2017, *24* (20), 1448–1453.

**Brox, Enzo and Daniel Goller**, "Tournaments, Contestant Heterogeneity and Performance," *Journal of Political Economy Microeconomics*, 2025, *3* (1), 83–126.

**Caeyers, Bet and Marcel Fafchamps**, "Exclusion Bias in the Estimation of Peer Effects," *Journal of Human Resources*, 2024, *59* (3), 757–796.

**Chowdhury, Subhasish M., Patricia Esteve-González, and Anwesha Mukherjee**, "Heterogeneity, Leveling the Playing Field, and Affirmative Action in Contests," *Southern Economic Journal*, 2023, *89* (3), 803–844.

**Coffey, Bentley and Michael T. Maloney**, "Beating a Live Horse: Effort's Marginal Cost Revealed in a Tournament," Working Paper, Clemson University, John E. Walker Department of Economics 2010.

**Drugov, Mikhail and Dmitry Ryvkin**, "Hunting for the Discouragement Effect in Contests," *Review of Economic Design*, 2025, *29* (1), 1–29.

**Emerson, John and Brian Hill**, "Peer Effects in Marathon Racing: The Role of Pace Setters," *Labour Economics*, 2018, *54*, 165–174.

**fei Lee, Lung**, "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, 2007, *140* (2), 333–374.

**Forrest, David**, "The Peculiar Economics of Horse Racing," in Brad R. Humphreys and Bernd Frick, eds., *Advances in Sports Economics*, Cambridge University Press, 2023.

**Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli**, "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–275.

**Guryan, Jonathan, Kris Kroft, and Matthew J. Notowidigdo**, "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 34–68.

**Hill, Brian**, "The Superstar Effect in 100-Meter Tournaments," *International Journal of Sport Finance*, 2014, *9* (2), 118–134.

**Lazear, Edward P. and Sherwin Rosen**, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

**Leung, Michael P.**, "Network Cluster-Robust Inference," *Econometrica*, 2023, *91* (2), 641–667.

**Manski, Charles F.**, "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

**Oster, Emily**, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

**Sacerdote, Bruce**, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

**Stock, James H. and Motohiro Yogo**, "Testing for Weak Instruments in Linear IV Regression," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 2005, pp. 80–108.

**Yamane, Shoko and Ryohei Hayashi**, "Peer Effects among Swimmers," *Scandinavian Journal of Economics*, 2015, *117* (4), 1230–1255.

# Appendix

## A. Data Description

Table A1: Data Feasibility Checks

| Check | Value | Threshold | Pass |
|---|---|---|---|
| Horses with 3+ starts | 43,534 | 2,000 | ✓ |
| Horse pairs with 2+ meetings | 134,645 | 500 | ✓ |
| Horses with 2+ different jockeys | 43,152 | 1,000 | ✓ |
| Claiming race share (%) | 60.1% | ∼30% | ✓ |
| Speed rating non-null (%) | 100.0% | >50% | ✓ |
| Pre-race quality coverage (%) | 98.6% | >40% | ✓ |
| Field size range | 2–18 (mean = 7.5) | Variation needed | ✓ |

*Note:* Each check confirms sufficient variation for the corresponding identification strategy.

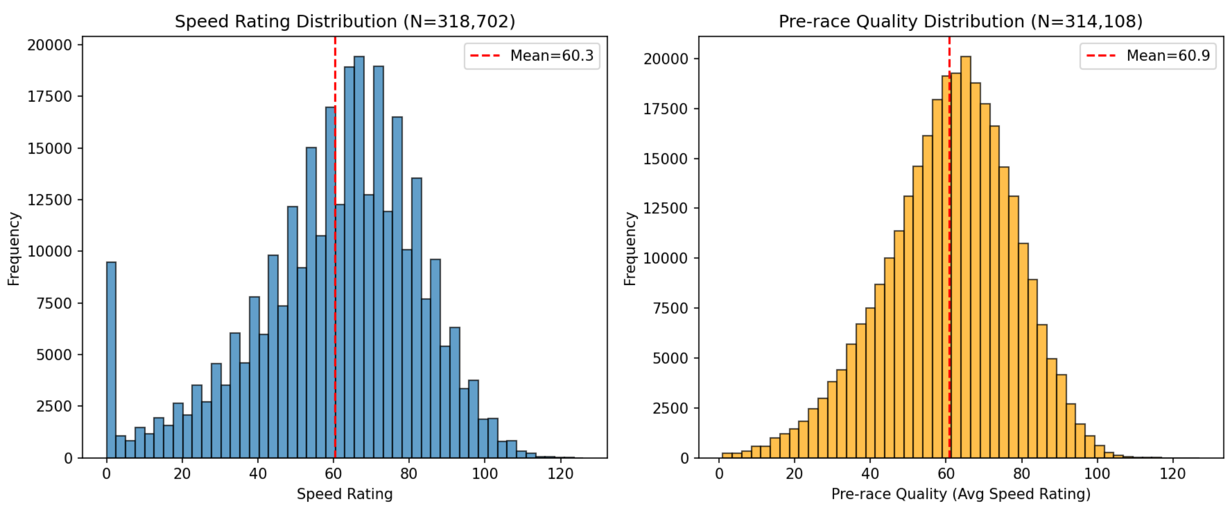


Figure A1: Distribution of Horse Quality (Speed Rating Residuals)

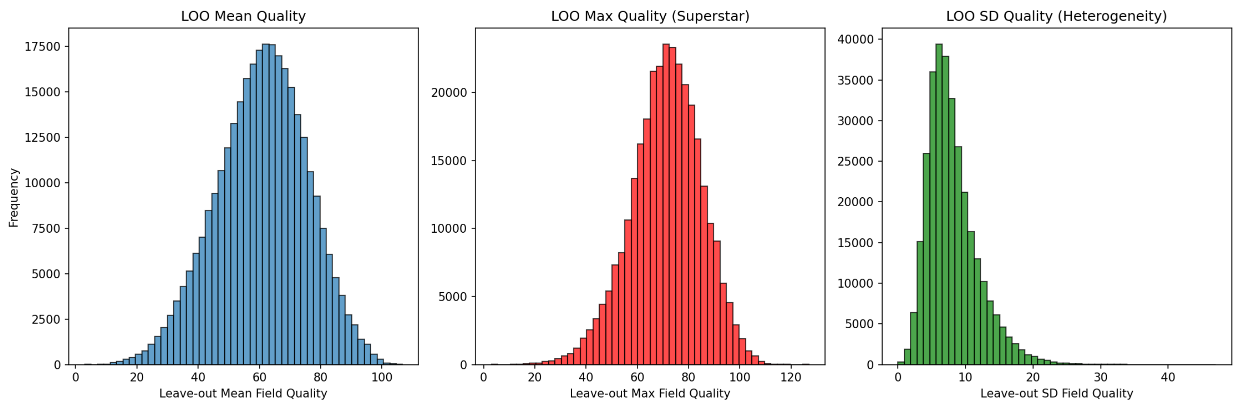


Figure A2: Field Composition by Race Type

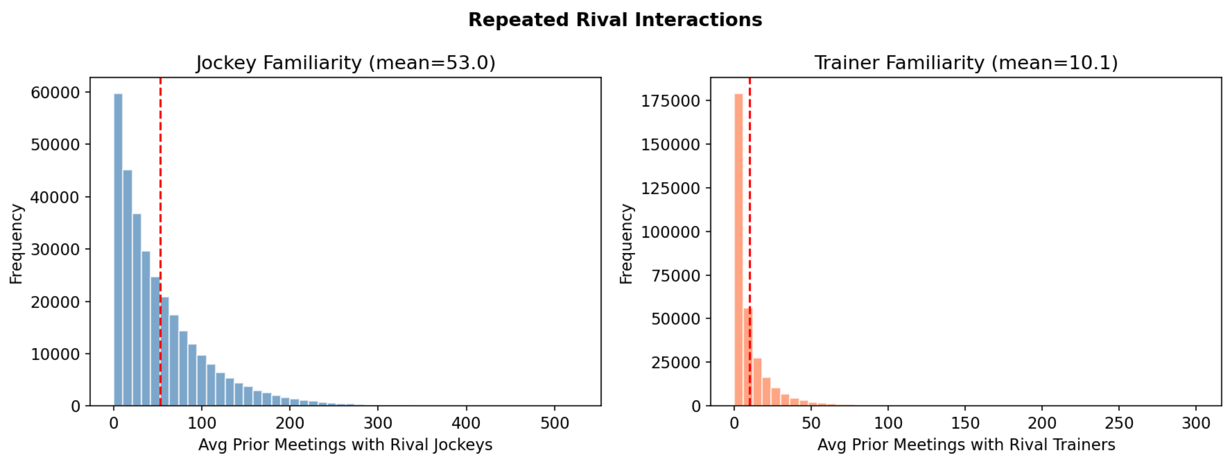


Figure A3: Distributions of Jockey and Trainer Rival Familiarity

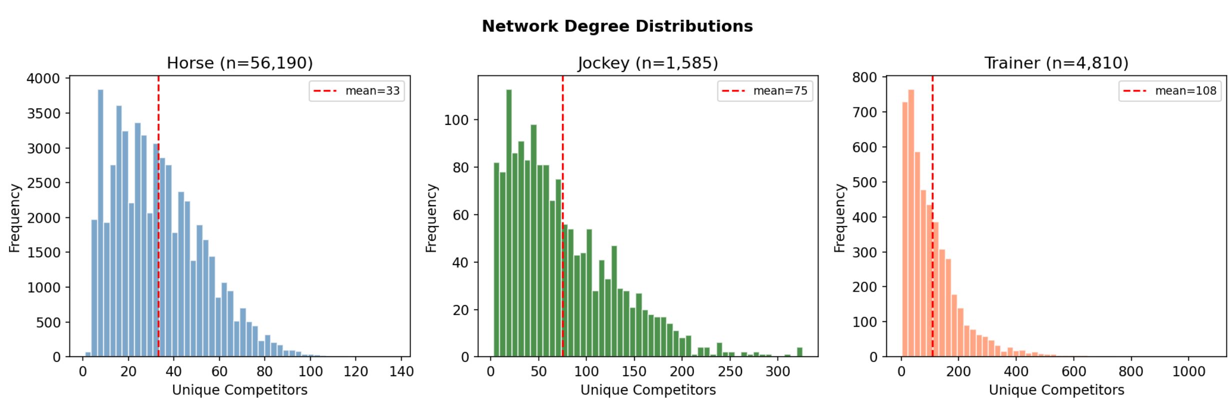


Figure A4: Network Degree Distributions for Horses, Jockeys, and Trainers

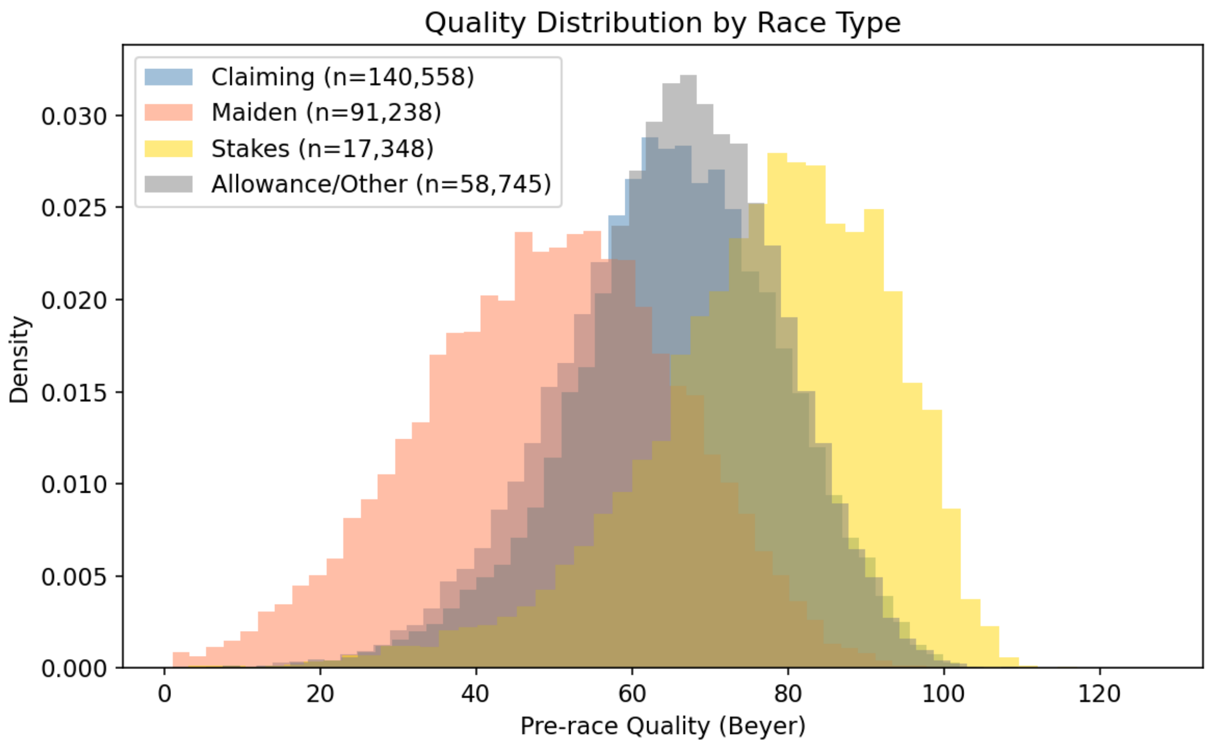


Figure A5: Quality Distributions by Race Type

## B. Additional Main Results

Table A2: Familiarity and Peer Effects

| | (1)<br>Baseline | (2)<br>Familiarity Levels | (3)<br>Familiarity Interactions |
|---|---|---|---|
| LOO Mean Field Quality | 0.188*** | 0.184*** | 0.209*** |
| | (0.00658) | (0.00654) | (0.00709) |
| LOO Mean Jockey Quality | -0.0851*** | -0.0910*** | -0.0915*** |
| | (0.0125) | (0.0124) | (0.0124) |
| LOO Mean Trainer Quality | -0.0631*** | -0.0634*** | -0.0656*** |
| | (0.00999) | (0.00999) | (0.00999) |
| Jockey Rival Familiarity (avg prior meetings) | | 0.0123*** | 0.0329*** |
| | | (0.000982) | (0.00366) |
| Trainer Rival Familiarity (avg prior meetings) | | 0.0350*** | 0.0945*** |
| | | (0.00336) | (0.0141) |
| Field Quality x Jockey Familiarity | | | -0.000310*** |
| | | | (0.0000517) |
| Field Quality x Trainer Familiarity | | | -0.000839*** |
| | | | (0.000194) |
| Observations | 304546 | 304546 | 304546 |

All specifications include horse, jockey, and trainer FE. SEs two-way clustered by race and horse.
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table A3: Heterogeneous Peer Effects: Interaction Analysis

| | (1)<br>Dirt | (2)<br>Sprint | (3)<br>Claiming | (4)<br>Maiden | (5)<br>Jockey Q | (6)<br>Familiarity | (7)<br>Experience |
|---|---|---|---|---|---|---|---|
| LOO Mean Field Quality | 0.107*** | 0.172*** | 0.181*** | 0.210*** | 0.147*** | 0.155*** | 0.166*** |
| | (0.00982) | (0.00578) | (0.00692) | (0.00616) | (0.00538) | (0.00534) | (0.00530) |
| Field Quality x Dirt Surface | 0.0228** | | | | | | |
| | (0.00914) | | | | | | |
| Field Quality x Sprint (<6f) | | -0.0443*** | | | | | |
| | | (0.00611) | | | | | |
| Field Quality x Claiming Race | | | -0.0353*** | | | | |
| | | | (0.00613) | | | | |
| Field Quality x Maiden Race | | | | -0.144*** | | | |
| | | | | (0.00874) | | | |
| Field Quality x Jockey Quality (std) | | | | | -0.00288 | | |
| | | | | | (0.00305) | | |
| Field Quality x Jockey Familiarity (std) | | | | | | -0.0231*** | |
| | | | | | | (0.00233) | |
| Field Quality x Prior Starts (std) | | | | | | | 0.0348*** |
| | | | | | | | (0.00438) |
| Observations | 311047 | 311047 | 311047 | 311047 | 309058 | 311047 | 311047 |

Horse FE in all specs. Continuous moderators standardized. Clustered by race.
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table A4: Field Size Sensitivity

| | (1) Full | (2) 4-14 | (3) 6-12 | (4) 8-14 |
|---|---|---|---|---|
| LOO Mean Field Quality | 0.156*** | 0.156*** | 0.163*** | 0.189*** |
| | (0.00537) | (0.00537) | (0.00579) | (0.00827) |
| Pre-race Quality (own) | -3.489*** | -3.486*** | -3.489*** | -3.423*** |
| | (0.0330) | (0.0331) | (0.0348) | (0.0468) |
| Observations | 311047 | 310535 | 281569 | 166234 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## C. Identification Diagnostics

Table A5: Placebo and Positive Control Tests

| | (1) Full | (2) Claiming | (3) Non-Claiming | (4) Lagged (+ ctrl) | (5) Joint |
|---|---|---|---|---|---|
| future_field_quality | 0.406*** | 0.484*** | 0.340*** | | 0.400*** |
| | (0.00452) | (0.00634) | (0.00716) | | (0.00448) |
| Pre-race Quality (own) | -4.705*** | -4.239*** | -5.004*** | -3.196*** | -4.794*** |
| | (0.0334) | (0.0493) | (0.0511) | (0.0388) | (0.0335) |
| lagged_field_quality | | | | 0.0798*** | |
| | | | | (0.00477) | |
| LOO Mean Field Quality | | | | | 0.120*** |
| | | | | | (0.00542) |
| Observations | 255407 | 150752 | 93697 | 255407 | 255407 |

Horse FE in all specs. Clustered by race.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Network-Clustered vs Race-Clustered Standard Errors

| | (1)<br>Horse FE (race) | (2)<br>Horse FE (network) | (3)<br>Joint (race) | (4)<br>Joint (network) |
|---|---|---|---|---|
| LOO Mean Field Quality | 0.156*** | 0.156*** | 0.188*** | 0.188*** |
| | (0.00537) | (0.0241) | (0.00658) | (0.0289) |
| Pre-race Quality (own) | -3.489*** | -3.489*** | -3.577*** | -3.577*** |
| | (0.0330) | (0.105) | (0.0334) | (0.0960) |
| LOO Mean Jockey Quality | | | -0.0851*** | -0.0851*** |
| | | | (0.0125) | (0.0172) |
| LOO Mean Trainer Quality | | | -0.0631*** | -0.0631*** |
| | | | (0.00999) | (0.0100) |
| Observations | 311047 | 311047 | 304546 | 304546 |

Cols 1-2: Spec 2 with race vs network clustering. Cols 3-4: Joint tripartite with race+horse vs network+horse clustering.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Anderson-Rubin Pre-Test for Peer Effects (Jung & Liu 2026)

| | (1)<br>LOO Mean | (2)<br>Mean+Max+SD | (3)<br>Top 5 Peers | (4)<br>Top 10 Peers | (5)<br>Mean + Peers |
|---|---|---|---|---|---|
| LOO Mean Field Quality | 0.156*** | 0.203*** | | | 0.266*** |
| | (0.00537) | (0.0140) | | | (0.0243) |
| LOO Max Field Quality (Superstar) | | -0.0492*** | | | |
| | | (0.0131) | | | |
| LOO SD Field Quality (Heterogeneity) | | 0.0306* | | | |
| | | (0.0179) | | | |
| peer_q_rank1 | | | 0.00804 | -0.00523 | -0.0330*** |
| | | | (0.00941) | (0.0498) | (0.0101) |
| peer_q_rank2 | | | 0.0134 | 0.0720 | -0.0301** |
| | | | (0.0139) | (0.0816) | (0.0145) |
| peer_q_rank3 | | | 0.0487*** | 0.0715 | -0.000863 |
| | | | (0.0149) | (0.0986) | (0.0156) |
| peer_q_rank4 | | | 0.0387*** | -0.167 | -0.0172 |
| | | | (0.0135) | (0.110) | (0.0145) |
| peer_q_rank5 | | | 0.0400*** | 0.183 | -0.0233** |
| | | | (0.00889) | (0.123) | (0.0105) |
| Observations | 311047 | 310893 | 279912 | 14341 | 279912 |

Horse FE in all specs. Clustered by race. H2a-H2b: ranked individual peer qualities as separate regressors. H3a: tests sufficiency of LOO mean.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Tripartite Anderson-Rubin Pre-Test

| | (1)<br>Ranked Peers Only | (2)<br>LOO Means + Ranked Peers |
|---|---|---|
| peer_q_rank1 | 0.0289*** | -0.00963 |
| | (0.00827) | (0.00870) |
| peer_q_rank2 | 0.0388*** | -0.00867 |
| | (0.0118) | (0.0123) |
| peer_q_rank3 | 0.0899*** | -0.0185 |
| | (0.00973) | (0.0117) |
| peer_jq_rank1 | -0.0560*** | -0.0546*** |
| | (0.0124) | (0.0128) |
| peer_jq_rank2 | 0.00634 | 0.00239 |
| | (0.0181) | (0.0187) |
| peer_jq_rank3 | -0.0372** | -0.0318* |
| | (0.0159) | (0.0191) |
| peer_tq_rank1 | -0.00783 | -0.00322 |
| | (0.00836) | (0.00888) |
| peer_tq_rank2 | -0.0206* | -0.0176 |
| | (0.0122) | (0.0126) |
| peer_tq_rank3 | -0.0256** | -0.0165 |
| | (0.0103) | (0.0122) |
| LOO Mean Field Quality | | 0.224*** |
| | | (0.0148) |
| LOO Mean Jockey Quality | | -0.0113 |
| | | (0.0229) |
| LOO Mean Trainer Quality | | -0.0324* |
| | | (0.0171) |
| Observations | 303243 | 303243 |

Horse+Jockey+Trainer FE. Two-way clustered by race and horse. Top 3 ranked peers per agent type.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Entry Model: Scratch Decisions and Field Quality

| | (1)<br>Any Scratch | (2)<br>Strategic (A) | (3)<br>Interaction | (4)<br>Claiming | (5)<br>Non-Claiming |
|---|---|---|---|---|---|
| main | | | | | |
| Expected LOO Field Quality | 0.00780*** | 0.0109*** | 0.0490*** | -0.00659*** | 0.0260*** |
| | (0.00157) | (0.00166) | (0.00322) | (0.00245) | (0.00233) |
| Own Quality (avg pre-race) | -0.00886*** | -0.00794*** | 0.0282*** | -0.00236 | -0.0114*** |
| | (0.00116) | (0.00130) | (0.00307) | (0.00209) | (0.00165) |
| Entered Field Size | 0.163*** | 0.178*** | 0.177*** | 0.181*** | 0.177*** |
| | (0.00210) | (0.00252) | (0.00253) | (0.00350) | (0.00373) |
| purse | -0.000000822*** | -0.000000649*** | -0.000000293** | -0.00000239*** | -0.000000474*** |
| | (0.000000122) | (0.000000134) | (0.000000125) | (0.000000843) | (0.000000132) |
| distance | 0.000431*** | -0.000460*** | -0.000424*** | 0.000172 | -0.000778*** |
| | (0.0000672) | (0.0000789) | (0.0000797) | (0.000135) | (0.000103) |
| Expected Field Quality x Own Quality | | | -0.000622*** | | |
| | | | (0.0000467) | | |
| Observations | 582192 | 582192 | 582192 | 346328 | 235864 |

Logit coefficients. SEs clustered by horse. E4 splits by claiming status.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Race-Level Scratch Intensity and Field Quality

| | (1)<br>All Scratches | (2)<br>A-type | (3)<br>A-type (Claiming) | (4)<br>A-type (Non-Claiming) |
|---|---|---|---|---|
| Race Mean Quality | 0.00154 | -0.00662*** | -0.00653*** | -0.00686*** |
| | (0.000960) | (0.000655) | (0.000806) | (0.00118) |
| (first) purse | -0.000000319 | -0.000000517*** | -0.000000839 | -0.000000641*** |
| | (0.000000282) | (0.000000142) | (0.000000844) | (0.000000159) |
| (first) field_size | -0.0793*** | 0.256*** | 0.217*** | 0.303*** |
| | (0.00846) | (0.00548) | (0.00657) | (0.00950) |
| Observations | 42613 | 23945 | 14667 | 9278 |

OLS at race level. Robust SEs. Track FE in all specs.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## D. Extensions

Table A11: Network Position and Peer Effects

| | (1) Controls | (2) Eigen x FQ | (3) Cluster x FQ | (4) PR x FQ | (5) Tripartite + Centrality |
|---|---|---|---|---|---|
| LOO Mean Field Quality | 0.156*** | 0.151*** | 0.294*** | -0.0137 | 0.188*** |
| | (0.00537) | (0.00537) | (0.0114) | (0.0120) | (0.00658) |
| Field Quality x Eigenvector Centrality | | 0.744*** | | | |
| | | (0.107) | | | |
| Field Quality x Clustering Coefficient | | | -0.567*** | | |
| | | | (0.0406) | | |
| Field Quality x PageRank | | | | 0.537*** | |
| | | | | (0.0348) | |
| LOO Mean Jockey Quality | | | | | -0.0851*** |
| | | | | | (0.0125) |
| LOO Mean Trainer Quality | | | | | -0.0631*** |
| | | | | | (0.00999) |
| Observations | 311047 | 311047 | 311047 | 311047 | 304546 |

Cols 1-4: Horse FE, clustered by race. Col 5: Horse+Jockey+Trainer FE, two-way clustered.

Centrality levels absorbed by horse (cols 1-4) and agent (col 5) fixed effects; interactions reported where non-collinear.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A12: Temporal Dynamics: First vs Second Half of Year

| | (1)<br>H1 Horse FE | (2)<br>H2 Horse FE | (3)<br>H1 Tripartite | (4)<br>H2 Tripartite | (5)<br>Pooled Interaction |
|---|---|---|---|---|---|
| LOO Mean Field Quality | 0.0871*** | 0.0990*** | 0.112*** | 0.142*** | 0.203*** |
| | (0.00790) | (0.00779) | (0.00938) | (0.00924) | (0.00602) |
| Pre-race Quality (own) | -5.729*** | -3.590*** | -5.791*** | -3.635*** | -3.529*** |
| | (0.0565) | (0.0498) | (0.0618) | (0.0504) | (0.0332) |
| LOO Mean Jockey Quality | | | -0.0582*** | -0.0982*** | |
| | | | (0.0166) | (0.0188) | |
| LOO Mean Trainer Quality | | | -0.0517*** | -0.0489*** | |
| | | | (0.0127) | (0.0157) | |
| Field Quality x H2 | | | | | -0.0659*** |
| | | | | | (0.00506) |
| Second Half of Year | | | | | 5.723*** |
| | | | | | (0.326) |
| Observations | 139305 | 162221 | 133484 | 161145 | 311047 |

Cols 1-2: Horse FE, clustered by race. Cols 3-4: Full tripartite, two-way clustered. Col 5: Pooled interaction test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Jockey Switching Quasi-Experiment

| | (1)<br>Levels | (2)<br>Levels+Fam | (3)<br>FD | (4)<br>FD+Controls | (5)<br>Falsification | (6)<br>Interaction |
|---|---|---|---|---|---|---|
| Delta Jockey Quality (new - old) | 0.0413*** | 0.0340*** | 0.0159** | 0.0219*** | 0.138*** | 0.0767*** |
| | (0.00462) | (0.00470) | (0.00710) | (0.00717) | (0.0194) | (0.0196) |
| Pre-race Quality (own) | -3.567*** | -3.600*** | | | -2.800*** | -3.704*** |
| | (0.0582) | (0.0585) | | | (0.0627) | (0.0587) |
| Jockey Rival Familiarity (avg prior meetings) | | 0.00868*** | | -0.00574*** | | |
| | | (0.000988) | | (0.000829) | | |
| Days Since Last Race | | | | 0.00451*** | | |
| | | | | (0.00175) | | |
| LOO Mean Field Quality | | | | | | 0.161*** |
| | | | | | | (0.00843) |
| Delta Jockey Quality x Field Quality | | | | | | -0.000655** |
| | | | | | | (0.000312) |
| Observations | 128606 | 128606 | 138804 | 138804 | 108799 | 128606 |

G1-G2, G5-G6: Horse FE, clustered by horse. G3-G4: First-difference, clustered by horse. G5: Non-switch subsample (falsification).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## E. Advanced Methods

Table A14: Causal Forest Heterogeneous Treatment Effects (Claiming Subsample)

| *Panel A: Treatment Effect Estimates* | | | | |
|---|---|---|---|---|
| Specification | ATE | SE | $N$ | |
| Demeaned (baseline) | 0.288*** | (0.006) | 179,435 | |
| No demeaning | 0.194*** | (0.021) | 179,435 | |
| Frequent racers ($\geq$ 5 races) | 0.314*** | (0.007) | 136,260 | |
| *Calibration Diagnostics* | | | | |
| | Estimate | SE | $p$-value | |
| Mean forest prediction | 0.992 | (0.020) | | should $\approx$ 1 |
| Differential forest prediction | 1.332 | (0.045) | <0.001 | rejects homogeneity |

| *Panel B: Variable Importance (Top 10, Demeaned Specification)* | | |
|---|---|---|
| Rank | Variable | Importance |
| 1 | Horse clustering coefficient | 0.208 |
| 2 | Horse eigenvector centrality | 0.151 |
| 3 | Distance | 0.128 |
| 4 | Pre-race quality (own) | 0.116 |
| 5 | Field size | 0.100 |
| 6 | Pre-race starts | 0.064 |
| 7 | Field quality LOO max | 0.063 |
| 8 | Purse | 0.037 |
| 9 | Jockey–rival familiarity | 0.024 |
| 10 | Trainer–rival familiarity | 0.018 |

*Note:* Estimated using the `grf` package. Treatment is leave-out mean field quality. Horse fixed effects absorbed via residualization.
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

Table A15: Convexity and Supermodularity Tests

| *Panel A: Polynomial Specification* | | | | |
|---|---|---|---|---|
| Term | Estimate | SE | $p$-value | Interpretation |
| Field quality (linear) | 0.221*** | (0.005) | | |
| Field quality (squared) | 0.001** | (0.000) | 0.013 | Convex |
| Field quality × own quality | 0.005*** | (0.000) | <0.001 | Supermodular |
| *Panel B: Generalized Additive Model* | | | | |
| Term | EDF | | $p$-value | Interpretation |
| Tensor product: field quality, own quality | 16.3 | | <0.001 | Significant nonlinear interaction |
| *Panel C: Decile-Specific Peer Effects* | | | | |
| Decile | Estimate | SE | Mean Quality | |
| D1 | 0.106*** | (0.015) | 28.6 | |
| D2 | 0.182*** | (0.015) | 43.0 | |
| D3 | 0.194*** | (0.013) | 50.3 | |
| D4 | 0.241*** | (0.013) | 55.7 | |
| D5 | 0.277*** | (0.014) | 60.1 | |
| D6 | 0.239*** | (0.014) | 64.2 | |
| D7 | 0.270*** | (0.013) | 68.3 | |
| D8 | 0.283*** | (0.013) | 72.8 | |
| D9 | 0.333*** | (0.014) | 78.2 | |
| D10 | 0.310*** | (0.016) | 87.9 | |

*Note:* All specifications include horse fixed effects. $N = 314{,}079$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Partial Network Bias Correction

| *Panel A: Analytical Correction ($\hat{\beta}/\rho$)* | | | |
|---|---|---|---|
| $\rho$ | Corrected Estimate | | |
| 0.1 | 2.045 | | |
| 0.3 | 0.682 | | |
| 0.5 | 0.409 | | |
| 0.7 | 0.292 | | |
| 1.0 | 0.205 | | (observed) |
| *Panel B: Simulation-Based Correction (50 Simulations)* | | | |
| $\rho$ | Corrected Estimate | Attenuation | |
| 0.1 | 9.489 | 2.2% | |
| 0.3 | 2.548 | 8.0% | |
| 0.5 | 1.181 | 17.3% | |
| 0.7 | 0.591 | 34.6% | |
| 1.0 | 0.205 | — | (no correction) |

*Note:* $\rho$ denotes the fraction of the full competition network observed. Our one-year sample corresponds to roughly $\rho = 0.3$. Observed estimates are lower bounds.

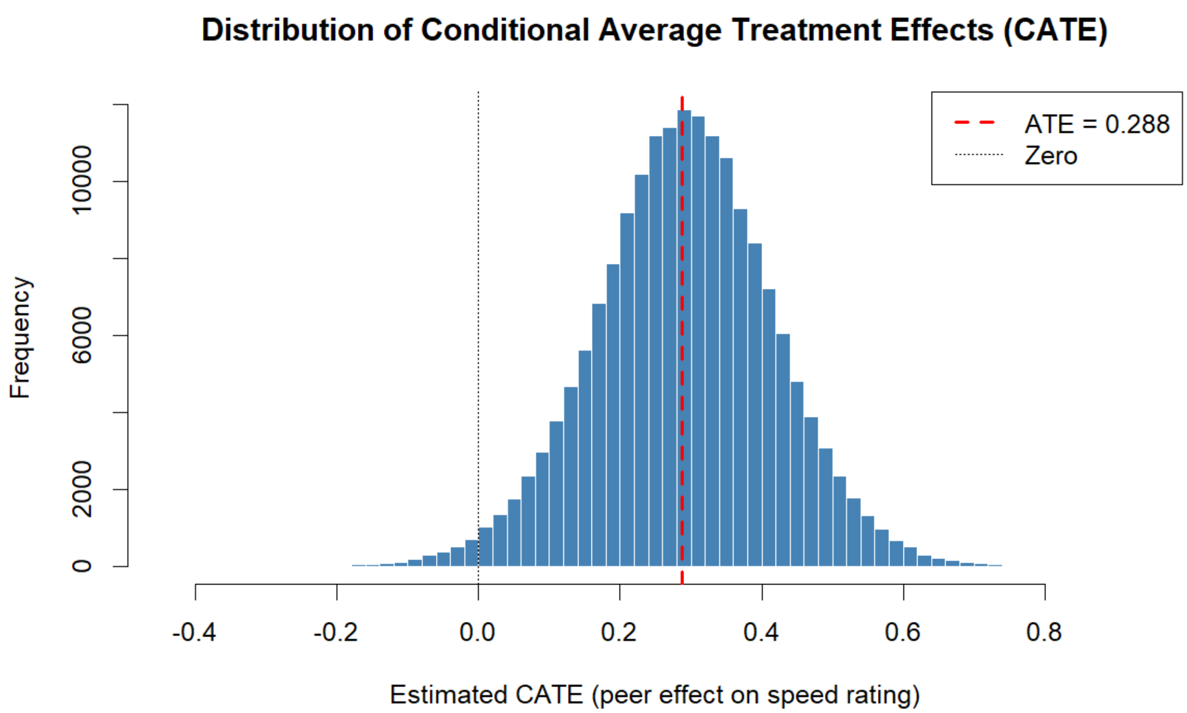


Figure A6: Distribution of Conditional Average Treatment Effects from Causal Forest

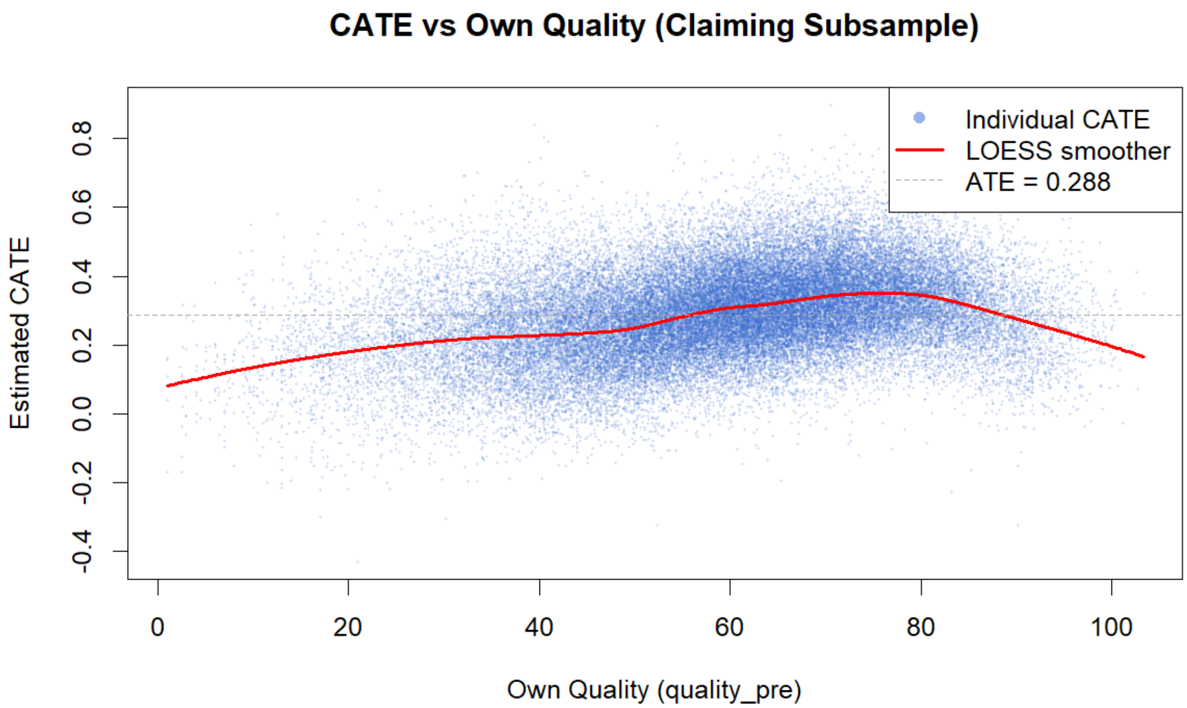


Figure A7: Causal Forest Conditional Average Treatment Effect (CATE) Estimates by Horse Quality Decile

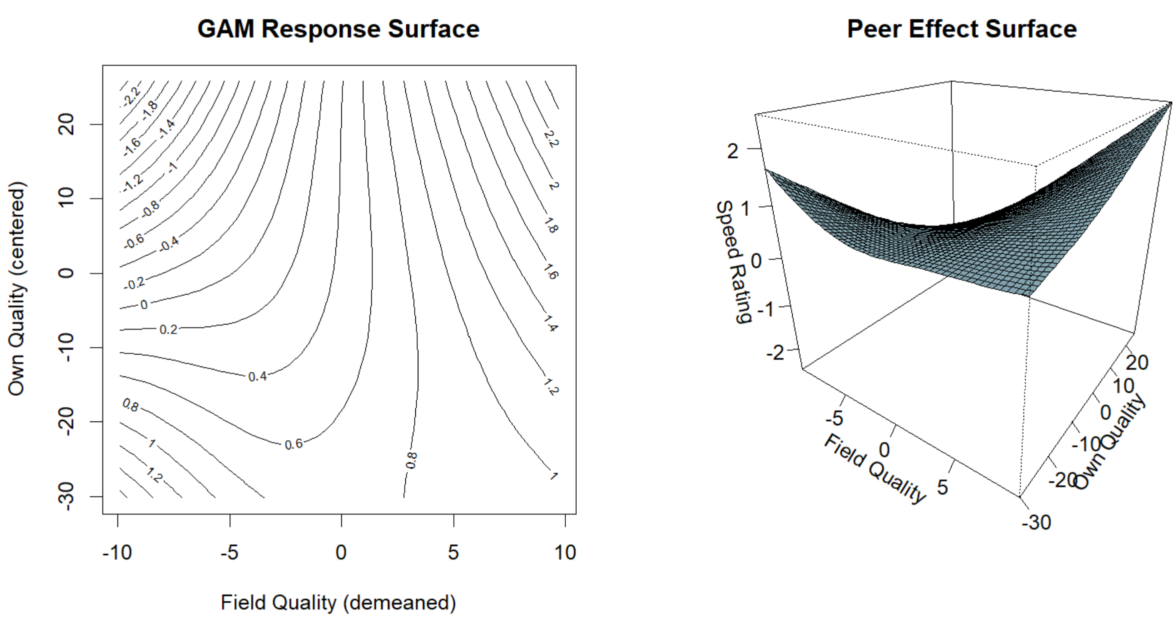


Figure A8: Convexity of Peer Effects: Response Surface over Own and Field Quality

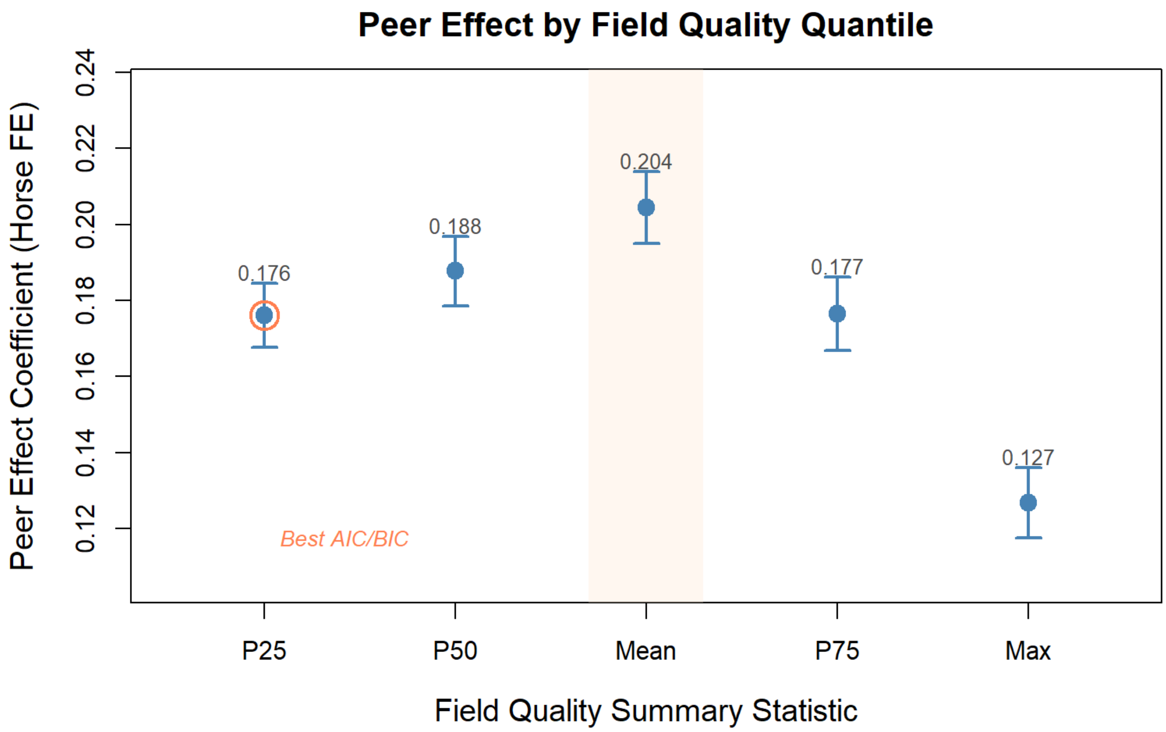


Figure A9: Peer Effect Coefficients across the Conditional Speed Rating Distribution

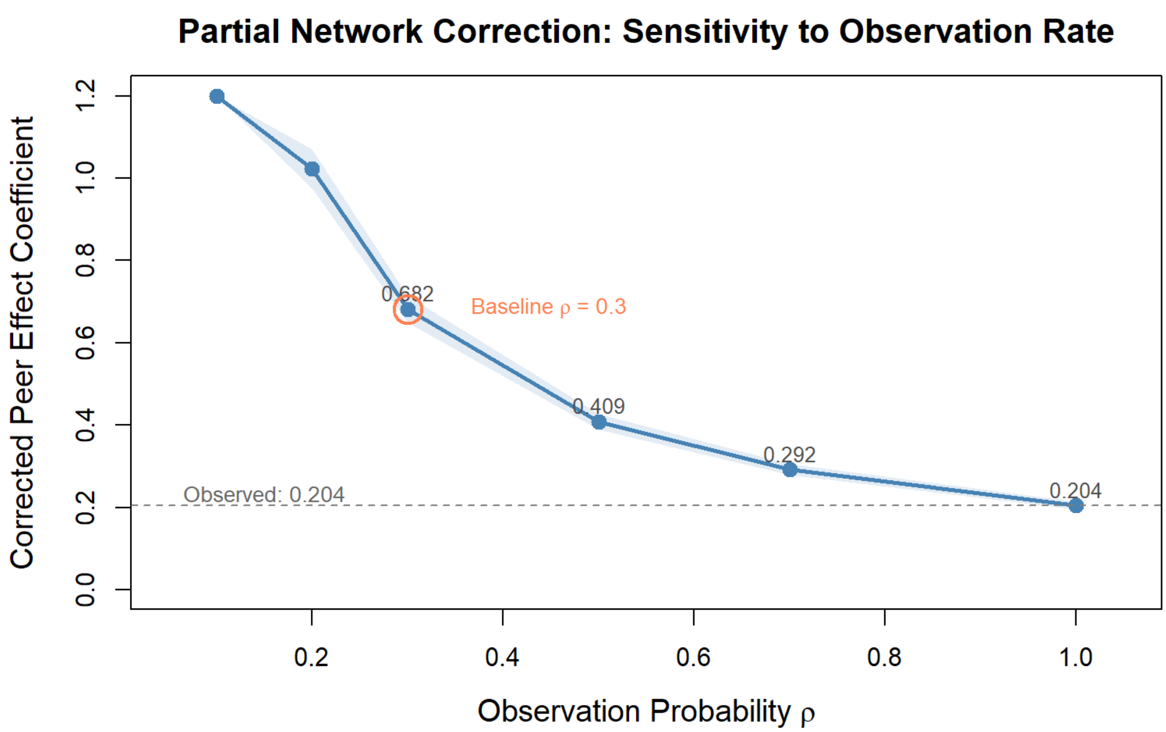


Figure A10: Sensitivity of Peer Effect Estimates to Partial Network Observation

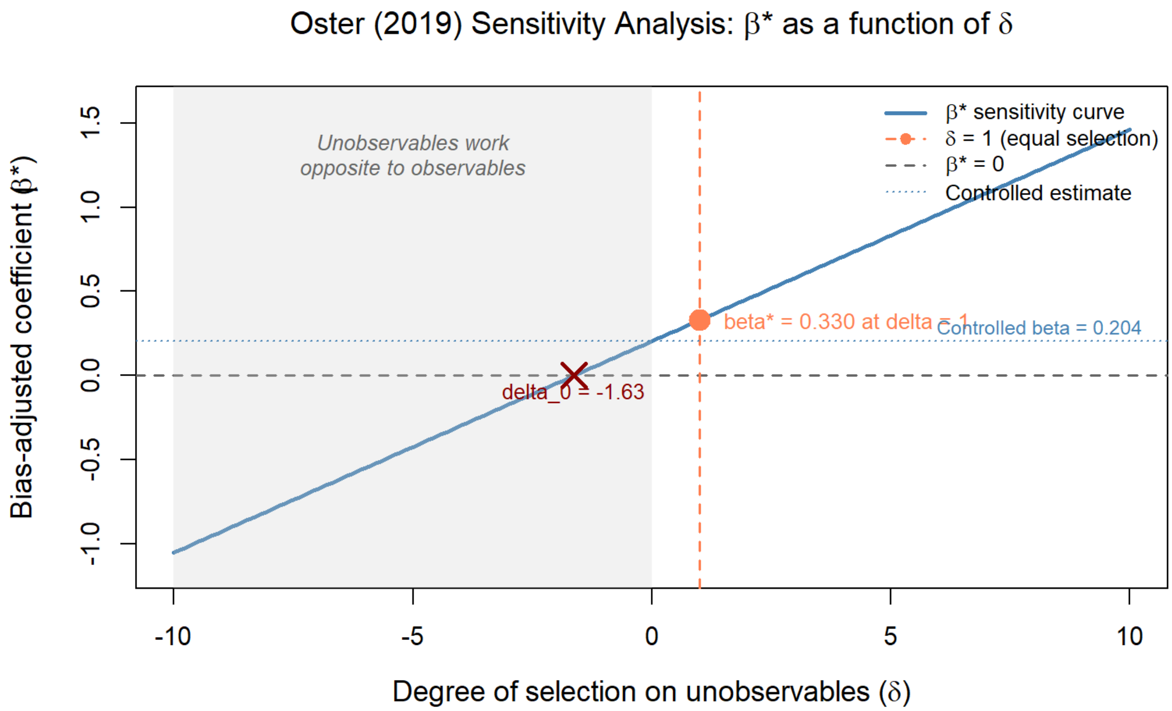


Figure A11: Oster (2019) Bounds on Peer Effect Estimates under Proportional Selection

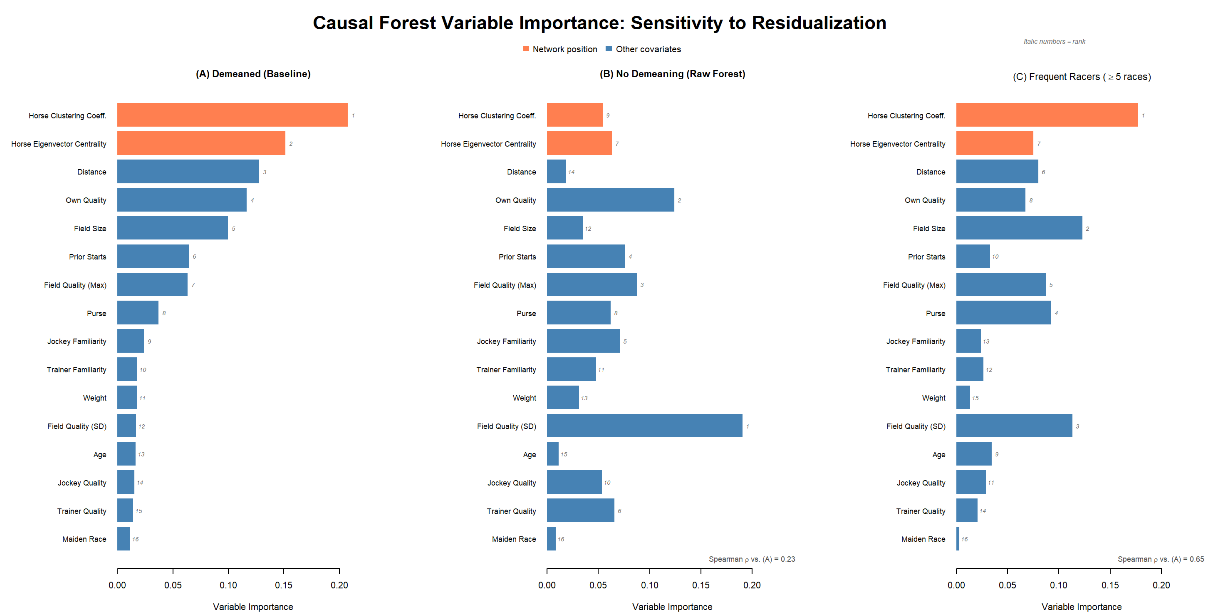


Figure A12: Variable Importance Comparison: Causal Forest vs. Gradient Boosting

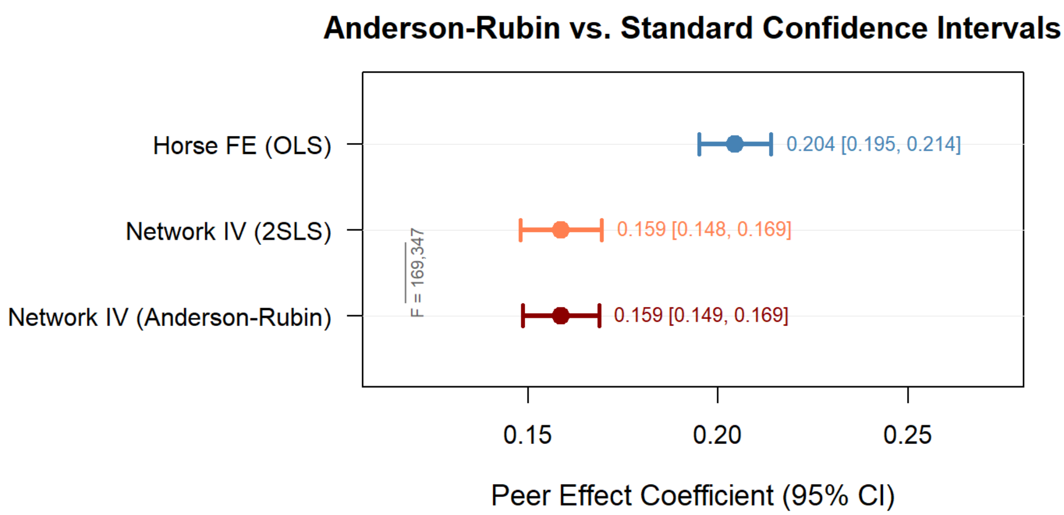


Figure A13: Anderson–Rubin Confidence Intervals vs. Conventional IV Inference