Peer Effects in Horse Racing:

A Tripartite Decomposition of Competitive Responses

Tandem Young*

Department of Economics

University of Arkansas

June 1, 2026

Abstract

Does the quality of one's competitors causally raise or depress individual performance? Credible answers are scarce because competitors are rarely assigned exogenously. We study the 2023 universe of North American thoroughbred racing (36,501 races; 261,496 starts), a setting that combines a standardized continuous output, within-horse variation in opponents, and institutional sources of plausibly exogenous variation in field composition (post-entry scratches and claiming-price constraints). Across pooled OLS, horse fixed effects, three scratch-based instruments, a claiming-race subsample, and a network instrument, a stronger field *raises* a horse's own speed rating: horses "rise to competition," reversing the discouragement effect documented among professional golfers. The central estimate is about 0.19 speed-rating points per unit of competitor quality—on the order of 1.4 horse lengths per standard-deviation shift in field quality—and the response is convex in own ability, though a pre-registered within-claim test supports this pattern only partially. The paper's main contribution is a tripartite decomposition of the peer effect into the three agents who jointly produce performance: the horse's competitive response is positive (+0.23), while jockeys and trainers respond negatively (−0.10 and −0.07), so the aggregate understates the horse's own response. A pre-registered placebo test does not pass, flagging selection into races; we decompose the resulting bias and report the magnitude as a bracket (roughly 0.10 to 0.48) rather than a single point.

JEL Codes: D82, J24, L83, Z20
Keywords: peer effects, tournament theory, horse racing, tripartite decomposition, instrumental variables

*Department of Economics, Sam M. Walton College of Business, University of Arkansas. Data provided by Equibase Company LLC. © 2026 Equibase Company LLC, all rights reserved. The full Equibase data disclaimer is reproduced in Appendix F.

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) organizes the identification challenge into three components: *endogenous effects*, through which peers' *outcomes* influence one's own; *contextual* or exogenous effects, through which peers' *characteristics* influence one's own outcome; and *correlated effects* arising from shared environmental shocks or selection into groups. Because we use peer characteristics (pre-race quality) rather than peer outcomes as the treatment, our target is the contextual effect, which is identified under weaker assumptions than the endogenous effect and does not require resolving Manski's reflection problem in its original formulation. The remaining threats to identification are correlated shocks and selection into races. Baseline entry is endogenous—horses sort into races by assessed value and connections' strategy—so we do not treat field composition itself as quasi-random. Instead, we exploit post-entry scratches that plausibly perturb the realized field after entries have closed, and we lean on consistency across strategies with distinct assumptions to address both threats.

This paper exploits this natural laboratory for studying peer effects in competitive settings. Horse racing offers four features that, taken together, provide a favorable identification environment. First, individual output is standardized and continuous: Equibase speed ratings 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, while baseline entry reflects sorting on assessed value, institutional features of the racing industry generate post-entry shocks to the realized field that are plausibly exogenous conditional on the entered roster: late off-the-turf and steward scratches remove competitors after entries have closed, and claiming price constraints restrict entry to horses of similar assessed value. A date-based preference system for oversubscribed races operates in the background, though we do not exploit it as a formal instrument. These features provide several complementary sources of identifying variation; we therefore lean on no single instrument 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 36,501 races and 261,496 individual entries. Identification rests on several strategies with distinct assumptions: pooled ordinary least squares (OLS) with extensive controls, horse fixed effects (FE), three instrumental variables exploiting post-entry scratches—veterinary, off-the-turf/steward, and also-eligible withdrawals that alter field composition after entries have closed—a claiming-race subsample where institutional constraints limit strategic entry, and a network IV derived from scheduling intransitivities. The fixed-effects specifications converge on roughly 0.15–0.19 speed-rating points per unit of competitor quality, and the instrumental-variables estimates bracket this band. We take the horse fixed-effects estimate of 0.193 as the central magnitude: it is stable across pooled OLS (0.148) and triple fixed effects (0.184) and across an extensive robustness battery. Two biases bound it in opposite directions. Classical measurement error in the generated peer mean attenuates the coefficient downward—a precision-stratified analysis recovers 0.246 in the most precisely measured fields, and an Oster bound that rises with controls ($\delta_0 = -1.79$) likewise places the true effect above the FE estimate. Against these, a horse fixed effects regression whose running-average own-quality control excludes the three most recent starts yields a conservative lower bound of roughly 0.10 (0.103); this neutralizes the Nickell (1981) dynamic-panel artifact in that control but also strips its purchase on contemporaneous form, so we read it as a lower bound rather than the preferred magnitude. The network instrument supplies a further floor of 0.128. The instrumental variables, which instrument the peer mean directly and do not identify the peer effect from the own-quality control, adjudicate between these biases and land in the upper part of the range: the veterinary-scratch instrument—the cleanest exogenous removal, free of the surface change that accompanies off-the-turf scratches—yields 0.257 on a sample five times larger than the off-the-turf design and is only modestly attenuated by surface fixed effects (0.226), while the off-the-turf/steward (0.484) and also-eligible (0.335) instruments bracket it. The scratch designs share a post-treatment bad-control channel. A pre-registered placebo test does not pass its committed thresholds—a within-claim forward placebo returns 0.277 and satisfies none of the three pre-specified criteria—so after the bad-control and persistence decomposition in Section 8.5 we read the scratch magnitudes as upper-bound local average treatment effects rather than clean point estimates. We discuss the Angrist (2014) concern that network-based instruments can proxy for group-level quality pools, and present a variance decomposition of the network instrument in Section 6. Sign consistency across strategies grounded in different identifying assumptions is the central diagnostic (see Figure 3). In economic terms, a one-standard-deviation increase in field quality (approximately 15.0 speed-rating points) translates to about 1.4 horse lengths at the central estimate—a meaningful margin in a sport where races are routinely decided by a length or less. Taken together, these strategies do not pin a single point but bracket the causal magnitude in the range of roughly 0.10 to 0.48, with the fixed-effects estimate near 0.19 as our central value. Section 8 reports the full range across specifications.

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 (+0.226 speed-rating 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 (−0.098 and −0.068, respectively), consistent with effort reallocation or conservative positioning when facing superior competition. The central estimate is therefore a net magnitude: it folds a larger positive horse response—the +0.226 horse channel exceeds the central horse-FE estimate of 0.193—against negative agent channels, so the aggregate understates the competitive response of the horse itself. 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). A pre-registered within-claim test supports this pattern only partially (a within-claim quartile range of 0.069), so a role for between-class sorting in the full-sample convexity cannot be ruled out (Section 7.4).

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 develops a Lazear–Rosen tournament framework with multi-position prizes that organizes the empirical findings: the convex positive peer response, the contrast with Brown (2011)'s superstar discouragement result, and the sign heterogeneity in the tripartite decomposition. Section 4 reviews the relevant literatures on tournament theory, peer effects, and horse racing economics. Section 5 describes the data and key variable construction. Section 6 develops the empirical strategy, detailing each of the five identification approaches. Section 7 presents the main results and the tripartite decomposition. Section 8 reports extensive robustness checks, including sensitivity to functional form, alternative peer quality measures, placebo tests, attrition and balance checks, and within-quarter and within-region replications. Section 9 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 60 percent of all U.S. races—and roughly 67 percent 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.

Figure 1 illustrates the claiming-price quality-cap mechanism empirically. Across the observed range of claiming prices in 2023, both own pre-race quality and the leave-out mean of opposing horses' quality rise together with substantial overlap in the within-band quality range. The visual makes plain why the claiming-price institution delivers a tighter peer-quality distribution than other race types and why we treat the claiming subsample as the cleanest within-class identification setting in the empirical analysis below.

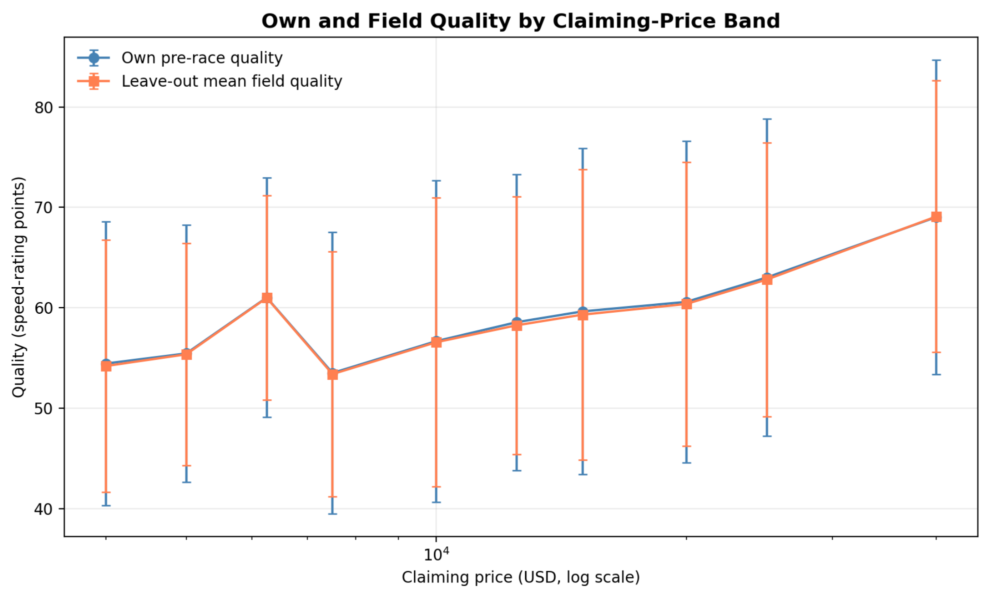


Figure 1: Own pre-race quality and leave-out mean field quality, by claiming-price band. Both rise together as the claiming price increases, with substantial within-band overlap, illustrating the institutional quality cap that motivates the claiming-subsample identification strategy. Whiskers are within-band standard deviations.

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. The Equibase data record a reason code for each scratch, which we group into five categories. *Trainer or owner scratches* are voluntary withdrawals that may reflect private information about a horse's fitness, making them potentially endogenous. *Off-the-turf scratches* occur when a race is moved off the turf course—typically because of rain—and horses entered specifically for the turf are withdrawn. *Steward scratches* are imposed by racing officials for administrative or regulatory reasons. *Veterinarian scratches* are mandated by the regulatory veterinarian when a horse is deemed unfit to compete on race day, for reasons of soundness or safety that are independent of the remaining horses' fitness.[1] *Also-eligible* entries are horses on an alternate list who fill vacancies created by scratches in the order drawn, so their appearance in the running field reflects the priority queue rather than trainer information about the remaining competitors. The off-the-turf and steward categories alter field composition after the entry decision has been made, for reasons—a weather-driven surface change or an administrative ruling about *another* horse—that are plausibly independent of the remaining horses' race-day fitness. This post-entry variation in field quality 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 6; because an off-the-turf scratch coincides with a surface change that could affect all runners directly, Section 6 reports a specification absorbing track×surface×distance fixed effects to hold that channel constant. A related source of quasi-random variation is the date-based preference rule that allocates starting positions when races are oversubscribed, again for reasons unrelated to trainer information about the remaining field.

Performance Measurement

The outcome variable throughout this paper is the Equibase speed rating,[2] a standardized performance metric that scales raw finishing times relative to typical times at each track and distance. On this scale a rating near 100 corresponds to track-record-caliber performance, 80 is strong, and 60

[1] The result charts record a "Veterinarian" scratch reason—15,333 such scratches in our data—and these form the basis of the veterinary instrument, the cleanest of our exogenous-withdrawal designs. Veterinary scrutiny carries regulatory force: under Horseracing Integrity and Safety Act (HISA) Rule 2262 the regulatory veterinarian holds "unconditional authority" to scratch a horse deemed unfit to compete (Horseracing Integrity and Safety Authority, Racetrack Safety Rule 2262; full regulations at `https://hisaus.org/regulations`). Because the veterinarian's determination turns on the scratched horse's own race-day soundness rather than on the trainer's private information about the rest of the field, it shifts the surviving field's composition for reasons plausibly orthogonal to the remaining runners' fitness.

[2] We use the speed figure distributed in the free Equibase dataset—the `SPEED_RATING` field of the result charts. This is Equibase's own track- and distance-adjusted figure, anchored to track-record benchmarks; it is distinct from the proprietary Beyer Speed Figure published by the Daily Racing Form, which is not contained in the Equibase data. We make no claim that the two are equivalent.

is roughly average; in our data only about 1.5 percent of runs exceed 100. One speed-rating point corresponds to approximately 0.50 horse lengths at the finish (about half a length; 0.42 in sprints to 0.58 in routes), so the peer effects estimated in this paper translate to physically meaningful shifts in competitive positioning. Because the figure is anchored to track-and-distance benchmarks, it provides a continuous outcome measure suitable for regression analysis, but it is anchored to each track's own historical times rather than placed on a single nationally comparable scale, so cross-track comparisons should be read with that caveat in mind. The free dataset does not document whether the figure nets out day-to-day variation in track speed (as a daily-track-variant adjustment would). This matters for the interpretation of common within-race shocks: an unusually fast track on race day raises all runners' contemporaneous ratings together. Because our treatment is the *backward-looking* mean of competitors' prior ratings, such a contemporaneous shock does not enter the peer measure directly, though it adds noise to the outcome. A raw-time replication that would sidestep the figure's adjustment entirely is infeasible here: the free dataset records an individual finishing time only for the small subset of horses (essentially race winners), and reconstructing times from the winner's clock and beaten-lengths margins would conflate absolute speed with relative finishing position.

Terminology

Several racing-specific terms appear throughout the analysis and merit brief definition.[3]

3 Theoretical Framework

This section develops a Lazear–Rosen tournament environment with heterogeneous types and multi-position prizes as the scaffold for the empirical exercise. The model organizes the parameter region in which the convex positive peer response documented in Section 7.4 arises in equilibrium and

[3] *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 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.

embeds the tripartite decomposition by allowing latent ability to be additive in three agent contributions with distinct effort technologies, providing a structural microfoundation for the channel-specific signs reported in Section 7.3. The model is intentionally minimal: its purpose is to discipline the empirical claims and locate the documented patterns in the existing theoretical literature, not to deliver a separate structural estimation.

3.1 Tournament Environment

A race is a Lazear–Rosen tournament with N heterogeneous entrants (Lazear and Rosen, 1981). Horse $i \in \{1, \dots, N\}$ chooses non-negative effort e_i to maximize expected utility from prize money minus an effort cost:

$$U_i = \sum_{k=1}^{K} \pi_k \, P\big(\text{rank}_i = k \mid e_i, e_{-i}\big) - c(e_i), \tag{1}$$

where π_k is the purse for finishing position k and $c(\cdot)$ is twice continuously differentiable, strictly increasing, and strictly convex with $c(0) = 0$. Performance follows the additive specification $y_i = a_i + e_i + \varepsilon_i$, where $a_i \sim F$ is horse i's latent ability and $\varepsilon_i \sim G$ is an i.i.d. idiosyncratic shock that captures race-day randomness. Each horse observes opponent abilities $\{a_j\}_{j \neq i}$ before choosing effort, but not opponent effort choices or shock realizations. The Bayes–Nash equilibrium is characterized by the first-order condition $c'(e_i^*) = \sum_{k=1}^{K} \pi_k \, \partial P(\text{rank}_i = k)/\partial e_i$, equating the marginal cost of effort to the expected marginal prize from a small effort increase.

3.2 Equilibrium Peer Response

The equilibrium peer response $\partial e_i^*/\partial \bar{a}_{-i}$ is determined jointly by the prize profile $\{\pi_k\}_{k=1}^{K}$, the curvature of the cost function $c(\cdot)$, and the curvature of the shock distribution $G(\cdot)$. Boucher et al. (2024) characterizes the parameter region in which the response is positive, convex in own ability, and consistent with a "rising to competition" pattern.

The canonical winner-takes-all tournament with Type-I extreme-value shocks ($\pi_1 = \pi$, $\pi_k = 0$ for $k \geq 2$) yields the closed-form equilibrium $e_i^* = \kappa \pi p_i^* (1 - p_i^*)$, where p_i^* is the equilibrium win probability and κ inversely indexes the marginal cost slope. This benchmark predicts a *discouragement* response in the typical region $p_i^* < \frac{1}{2}$: an increase in mean opponent ability lowers p_i^* further from the effort-maximizing point at $p_i^* = \frac{1}{2}$, reducing equilibrium effort. This is the parameter region documented by Brown (2011) in professional golf, where Tiger Woods's presence depressed non-superstar effort.

The empirical setting departs from this benchmark in two respects that reverse the prediction. First, North American thoroughbred purses are paid to the top five finishing positions (typically 60–20–10–5–3 percent of total), so the relevant marginal benefit of effort is the gradient $\sum_{k=1}^{K} \pi_k \, \partial P(\text{rank}_i = k)/\partial e_i$ rather than the win-probability gradient alone. Under multi-position prizes, the equilibrium effort response to opponent ability is positive across a substantially wider range of a_i than in the winner-takes-all case, because effort returns persist for non-leading positions (place, show, fourth, fifth). Second, racing involves production-side interaction through pace and drafting dynamics that the abstract tournament model does not capture; the front-runner specifications in Section 8.6 confirm that pace mechanics contribute to the positive aggregate response, while the peer effect among sole leaders (+0.100)—the smallest of the four running-style estimates, consistent with sole leaders being the horses least able to draft—remains positive and thereby establishes that the response is not solely a pace-following artifact.

The convexity in own ability documented in Section 7.4—the peer-response coefficient rises monotonically across own-quality quartiles from 0.229 to 0.328—is consistent with two distinct channels that the empirical design must separate. The first is compositional and operates as a selection threat rather than a behavioral mechanism: higher-ability horses systematically enter races with higher mean opponent quality (assortative entry within stakes and allowance categories), so part of the observed gradient could reflect between-class sorting rather than a within-horse response to a stronger field. The second is behavioral: the marginal effort productivity is itself increasing in own ability under the empirical prize profile, because higher-ability horses face fewer non-trivial opponents in any given race and therefore extract more marginal expected purse from each unit of effort, placing them in the parameter region where the multi-position effort gradient is steepest. Because the first channel is a confound rather than the mechanism of interest, the within-claim heterogeneity test of Section 7.4—which re-estimates the gradient across quartiles defined inside the institutionally constrained claiming band, holding between-class sorting roughly fixed—is what isolates the behavioral channel, and it yields only partial support. Boucher et al. (2024) formalizes the parameter conditions under which the convex positive pattern obtains; the empirical findings of this paper fall within the predicted region.

3.3 Tripartite Decomposition

The model embeds the tripartite decomposition via the additive ability specification

$$a_i = a_i^H + \alpha_J \, a_{j(i)}^J + \alpha_T \, a_{t(i)}^T, \tag{2}$$

where $j(i)$ and $t(i)$ index the jockey and trainer assigned to horse i and $\alpha_J, \alpha_T \geq 0$ are agent-specific weights mapping latent jockey and trainer ability into the focal horse's effective performance index. Each agent solves a separate effort problem with a distinct cost structure and objective. The horse's problem is the unconstrained tournament problem in (1), generating an unambiguously positive peer response, $\beta_H > 0$. The jockey and trainer problems include additional cost margins that the horse does not face: the jockey bears physical injury risk from aggressive riding, and the trainer manages horse condition across multiple starts within the racing season.

Under sufficiently steep risk-aversion or intertemporal-preservation penalties, these agent-specific cost margins reverse the sign of the peer response on the agent channels. A jockey facing a stronger field encounters more crowded racing lanes and a higher conditional probability of being trapped or fouled at the rail; the optimal response under risk aversion is to concede position rather than ride aggressively against a clearly superior field, generating $\beta_J < 0$. A trainer facing a likely defeat allocates preparation effort across multiple starts: the marginal return to aggressive race-day preparation against a likely-victorious field is small relative to the cost of compromising the horse's condition for the next more winnable start, generating $\beta_T < 0$. The empirical pattern $\beta_H > 0$, $\beta_J < 0$, $\beta_T < 0$ documented in Section 7.3 is the predicted equilibrium configuration when both agent friction margins bind.

3.4 Predictions

The model generates three predictions that the empirical exercise tests:

(P1) *Aggregate peer effect is positive.* Under multi-position purses with the empirically observed prize profile and production-side complementarities, the equilibrium effort response to mean opponent ability is positive for typical horses. Tested by the horse-FE specification (4) reported in Section 7.1.

(P2) *Peer response is convex in own ability.* A steeper effort-productivity gradient under the empirical prize profile places higher-ability horses in the parameter region characterized by Boucher et al. (2024), while the coincident assortative entry of higher-ability horses into stronger fields is a selection channel the test must separate from this behavioral prediction. Tested by the heterogeneity specification reported in Section 7.4; the documented monotonic increase in the peer-response coefficient across own-quality quartiles falls within the model's predicted region, with the within-claim test isolating the behavioral component and yielding partial support.

(P3) *Tripartite signs reflect agent objective heterogeneity.* The horse channel is positive; the jockey and trainer channels are negative when the risk-aversion and intertemporal-preservation cost margins bind. Tested by the joint tripartite specification (6) reported in Section 7.3.

The model thus does not derive a single closed-form prediction but provides the structural language for the empirical exercise. The convex positive peer response, the agent-channel sign heterogeneity, and the contrast with Brown (2011)'s superstar discouragement result are jointly consistent with a tournament environment combining multi-position purses, production-side complementarities, and agent-specific friction margins. The reduced-form findings reported in Sections 7 and 8 adjudicate which of these conditions hold in the racing setting; this section provides the conceptual scaffold for that adjudication.

4 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.

4.1 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 earliest empirical test of tournament theory is Ehrenberg and Bognanno (1990), who exploit prize-spread variation in PGA Tour events to validate the Lazear–Rosen prediction that effort responds to the gap between prizes. The most influential test of how competitor *composition* (rather than prize structure) 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. Deutscher et al. (2023) extend this evidence to professional tennis, finding similar shadow effects from elite players. 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 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. Genakos and Pagliero (2012) document interim-rank effects in Olympic weightlifting, where dynamic tournament structure produces within-contest heterogeneity in effort and risk-taking. 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.

4.2 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. Outside athletics, foundational evidence on peer effects in academic settings comes from Hoxby (2000), who exploits within-school variation in cohort composition; Carrell et al. (2013) subsequently formalize the distinction between exogenous and endogenously formed peer groups—a distinction that mirrors the contrast in our setting between scratch-induced and entry-driven variation in field composition.

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.

4.3 Identification of Peer Effects

The econometric challenge of identifying causal peer effects has been a central concern since Manski (1993) introduced 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 paper targets the contextual effect—the effect of peers' pre-race quality on own performance—and leaves the endogenous effect to future work; identifying the latter, the feedback from peers' realized *outcomes*, would require both peers' outcomes on the right-hand side and a second exclusion restriction to break the resulting simultaneity, which our design does not provide. The remaining identification challenges are therefore correlated shocks and selection into groups, which we address through a combination of high-dimensional fixed effects, instrumental variables, and institutional features that generate plausibly exogenous variation in field composition.

Even for contextual peer effects, several identification strategies are informative. Bramoullé et al. (2009) show that intransitivity in social networks generates exclusion restrictions useful for instrumenting endogenous peer outcomes; we construct a network-based instrument from the racing schedule as a supplementary check on our primary scratch-IV strategy (see also De Giorgi et al., 2010). An alternative strategy due to Graham (2008) exploits conditional variance restrictions rather than network structure; Goldsmith-Pinkham and Imbens (2013) characterize the identification problem in finite networks and provide tests for the network-IV exclusion restriction—cautions reflected in our supplementary treatment of the network IV. Lee (2007) demonstrates that identification is achievable even without network structure when group sizes vary, a condition satisfied in our setting where field sizes range from 2 to 18 starters.

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 several complementary 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. Recent work by Jochmans (2023) extends peer-effects identification to settings with endogenous interaction networks, a concern that is partially attenuated in our setting by the institutional structure of race entry.

4.4 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 performance (the standardized, track- and distance-adjusted Equibase speed rating), 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 (the standardized, track- and distance-adjusted Equibase speed rating), substantial quasi-random variation in field composition (from late scratches—off-the-turf, steward, and veterinary withdrawals—claiming-price entry constraints, and the date-based preference and also-eligible system), and a natural multi-level structure (horses, jockeys, and trainers) that permits channel decomposition.

4.5 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 a layered identification strategy that combines horse fixed effects, a scratch-based instrumental variable exploiting post-entry off-the-turf and steward withdrawals as the primary source of exogenous variation, a claiming race quasi-experiment where institutional constraints limit strategic entry, and triple fixed effects absorbing jockey and trainer sorting, with a supplementary network-based instrument from scheduling intransitivities (Bramoullé et al., 2009) reported as a cross-validation check—to establish that field composition causally affects individual horse performance (the standardized, track- and distance-adjusted Equibase speed rating).

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).

5 Data

Data Source

This paper uses the Equibase Free Dataset for the 2023 calendar year.[4] 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.

[4]Data provided by Equibase Company LLC. © 2026 Equibase Company LLC, all rights reserved. Equibase's full data attribution and disclaimer are reproduced in Appendix F.

Sample Construction

The estimation sample comprises 261,496 horse-race entries across 36,501 races at 120 tracks. The 120 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 45,486 unique horses, of which 35,311 have three or more starts—sufficient for within-horse fixed effects estimation. The median horse in the sample has five starts during 2023 (mean = 5.9). Data feasibility checks for the panel—record completeness, scratch-rate stability, and field-size coverage by race type—are reported in Appendix Table A1.

A total of 118,064 unique horse-pairs meet at least twice during the sample year, enabling the dyadic analysis of repeated matchups. Additionally, 35,347 horses (78 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 67 percent of races are claiming races, the race type with the most constrained entry sorting (see Section 2; Appendix Figure A5 plots the quality distribution by race type). 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.

Several features of Table 1 merit discussion. Equibase speed ratings average 62.6 with a standard deviation (SD) of 20.6, confirming substantial performance variation both within and between horses (Appendix Figure A1 plots the distribution of speed-rating residuals). Pre-race quality, the running average of a horse's prior Equibase speed ratings, has a tighter distribution (SD = 16.7) reflecting the averaging, but retains ample cross-sectional and longitudinal variation. The leave-out

Table 1: Summary Statistics

	N	Mean	SD	Min	P25	Median	P75	Max
Panel A: Performance and Quality								
Speed rating (DV)	261496	62.6	20.6	1	50.0	65.0	77.0	126
Pre-race quality (running avg)	259083	62.1	16.7	1	51.5	63.3	73.9	127
Panel B: Field Composition (Leave-Out)								
LOO mean field quality	259048	61.9	15.0	3	52.1	62.8	72.4	107
LOO max field quality	259048	72.0	13.9	3	63.2	72.7	81.7	127
LOO SD field quality	258915	7.8	3.6	0	5.3	7.2	9.6	47
Field size	261496	7.9	1.9	2	6.0	8.0	9.0	18
Panel C: Race and Horse Characteristics								
Official finish position	261496	4.4	2.5	1	2.0	4.0	6.0	18
Purse ($000s)	261496	40.9	91.1	2	14.1	25.0	42.0	6000
Distance (furlongs)	261496	6.3	1.9	1	5.5	6.0	8.0	32
Weight carried (lbs)	261496	120.8	3.1	103	119.0	121.0	123.0	154
Age (years)	261496	4.3	1.6	2	3.0	4.0	5.0	13

Notes: Unit of observation is a horse-race entry. Speed rating is the Equibase speed rating, 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.

mean of field quality—the main treatment variable—has a mean of 61.9 and a standard deviation of 15.0 (Appendix Figure A2 reports the field-composition distribution by race type); 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 (61.9) and leave-out maximum (72.0) 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 7.8, indicating meaningful heterogeneity in rival quality within fields. Approximately 0.9 percent of entries lack a pre-race quality measure, primarily debut runners with no prior recorded Equibase speed ratings; 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 $40,900, median $25,000, maximum $6 million).

Key Variable Definitions

Outcome variable. The dependent variable throughout is the Equibase speed rating, described in Section 2.

Pre-race quality. Each horse's quality entering a race is measured as the running average of all prior Equibase speed ratings.[5] 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 Equibase speed ratings 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 (off-the-turf/weather and

[5] Own pre-race quality, the leave-out field-quality means, and the jockey and trainer quality measures are all running-average *estimates* of latent ability rather than directly observed regressors, with sampling error that declines in the number of prior starts (or mounts/entries) on which each average is based. As generated regressors they carry classical measurement error, which attenuates the leave-out peer coefficient toward zero; the finite-sample leave-out exclusion-bias correction of Caeyers and Fafchamps (2024) that we apply is small in our setting. The standard errors reported throughout do not adjust for this first-step estimation; a two-step or bootstrap correction is left for future work.

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 6.

Supplementary network instrument. Following Bramoullé et al. (2009), we construct an indirect competitor quality measure as a supplementary cross-validation 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 6 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 Equibase speed rating 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. Appendix Figure A3 plots the distributions of jockey and trainer rival familiarity (the average number of prior meetings between an agent and the field's other agents).

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 $\times$ surface $\times$ distance fixed effects, which absorb all time-invariant features of a particular racing configuration at a given venue.

6 Empirical Strategy

6.1 The Identification Challenge

Our target parameter is the contextual peer effect in the Manski (1993) taxonomy: the causal effect of competitors' pre-race quality (a peer *characteristic*) on own speed rating. Because we do not include peers' *outcomes* on the right-hand side, we do not attempt to separately identify the endogenous peer effect, and the reflection problem in its original simultaneity form does not arise. Two identification threats remain in this setting. First, *correlated effects*: horses sharing a race also share race-day conditions—weather, track surface quality, pace dynamics—that independently affect all runners. Second, *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. This paper addresses these threats through a layered identification strategy that progressively strengthens the causal claim—from fixed effects that absorb permanent heterogeneity, to a scratch-based instrumental variable that isolates post-entry exogenous variation in field composition, to a quasi-experimental claiming-race subsample where institutional constraints limit strategic entry, with a supplementary network-based instrument reported for cross-validation.

6.2 Baseline Specification: OLS and Fixed Effects

The baseline pooled OLS specification takes the form

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

where y_{ir} is the Equibase speed rating 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 Equibase speed ratings, as defined in Section 5); 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 β 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 β 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:

$$y_{ir} = \alpha_i + \beta \bar{q}_{-i,r} + \gamma q_{ir} + \mathbf{x}'_{ir}\boldsymbol{\delta} + \varepsilon_{ir} \tag{4}$$

where α_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—β will capture this time-varying sorting rather than a causal peer effect. The subsequent identification strategies are designed to break this remaining endogeneity.

6.3 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:

$$y_{ir} = \alpha_i + \mu_j + \tau_t + \beta \bar{q}_{-i,r} + \gamma q_{ir} + \mathbf{x}'_{ir}\boldsymbol{\delta} + \varepsilon_{ir} \tag{5}$$

where μ_j is a fixed effect for jockey j and τ_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 β 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:

$$y_{ir} = \alpha_i + \mu_j + \tau_t + \beta_H \bar{q}^H_{-i,r} + \beta_J \bar{q}^J_{-i,r} + \beta_T \bar{q}^T_{-i,r} + \gamma_H q^H_{ir} + \gamma_J q^J_{ir} + \gamma_T q^T_{ir} + \mathbf{x}'_{ir}\boldsymbol{\delta} + \varepsilon_{ir} \tag{6}$$

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 5 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 β_H, β_J, and β_T jointly answer a question that the aggregate specification cannot: does a horse perform differently when facing a field of better *horses* (β_H), better *jockeys* (β_J), or better *trainers* (β_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 8).

6.4 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

$$\bar{q}_{-i,r} = \alpha_i + \gamma q_{ir} + \pi_1 z^{\text{scratch}}_{1r} + \pi_2 z^{\text{scratch}}_{2r} + \mathbf{x}'_{ir}\boldsymbol{\delta} + \nu_{ir} \tag{7}$$

where z^{scratch}_{1r} is the average pre-race quality of exogenously scratched (off-the-turf or steward) horses in race r and z^{scratch}_{2r} denotes the counts of off-the-turf and steward scratches, entered as separate instruments. The second stage replaces $\bar{q}_{-i,r}$ with its predicted value $\hat{\bar{q}}_{-i,r}$ from (7) and recovers the IV estimate of β.

The exclusion restriction requires that off-the-turf and steward scratches affect remaining horses' performance only through the change in field composition. This is grounded in the institutional process of Section 2: a race is moved off the turf because of weather, and steward scratches reflect administrative rulings about the *scratched* horse—neither is a function of the remaining horses' fitness or race-day readiness, so the scratch shifts the quality of the field that horse i faces without conveying information about horse i's own condition. On the exogenous-scratch sample—races with at least one off-the-turf or steward withdrawal, comprising roughly 11,453 horse-race observations after absorbing horse fixed effects—the instruments are strong, with a Kleibergen–Paap (KP) first-stage F-statistic of 103.8. In the first stage, the average quality of the exogenously scratched horses (z_{1r}^{scratch}) loads *positively* and highly significantly on the realized leave-out field quality—it co-moves with the race's overall quality level—while the count of steward scratches enters negatively, each withdrawal removing a competitor from the field that horse i faces. Two features support the exclusion restriction. First, the overidentification (Hansen J) test does not reject, and once we absorb track$\times$surface$\times$distance fixed effects—holding constant the surface change that accompanies an off-the-turf scratch—the estimate is essentially unchanged (from 0.484 to 0.469) while the overidentification test is comfortably satisfied ($p = 0.72$); the instrument therefore does not operate through the surface shift itself. Second, instrumenting instead on *all* scratches—folding in strategic trainer withdrawals—roughly halves the estimate (to 0.251) but *fails* the overidentification test decisively ($p = 1.1 \times 10^{-7}$), precisely the signature of an invalid, endogenous instrument; restricting to off-the-turf and steward scratches removes that violation (Table 6). The same exclusion logic extends to two further withdrawal types, constructed analogously (the average quality and count of withdrawn horses entered as instruments): *veterinary* scratches—withdrawals a state veterinarian orders on medical grounds, the cleanest exogenous removal and, unlike off-the-turf moves, free of any surface change—and *also-eligible* horses whose late draw-in or exclusion shifts the field. The veterinary instrument, estimated on five times the off-the-turf sample, is among our cleanest design-based exogeneity arguments (Section 7.2), though its overidentification test is only borderline (Hansen J $p = 0.054$). Each scratch instrument's magnitude is a local average treatment effect for its specific complier population and, as the placebo analysis of Section 8.5 shows, is best read as an upper-range estimate; the central population magnitude remains the horse fixed-effects estimate of 0.193, with the positive fixed-baseline and lagged-dependent dynamic-panel specifications that purge recent-form persistence returning roughly 0.10 as a conservative lower bound (the difference- and system-GMM estimators return negative, and hence unreliable, coefficients).

6.5 Supplementary Instrumental Variable: Competition Network

As a supplementary cross-validation of the scratch IV, we exploit the structure of the competition network following Bramoullé et al. (2009). The first-stage equation takes the form

$$\bar{q}_{-i,r} = \alpha_i + \gamma q_{ir} + \pi z_{ir}^{\text{net}} + \mathbf{x}_{ir}'\boldsymbol{\delta} + \nu_{ir} \tag{8}$$

where z_{ir}^{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. Appendix Figure A4 reports degree distributions for the horse, jockey, and trainer competition networks.

The network instrument is extremely strong, with a Kleibergen–Paap first-stage F-statistic of 144,566, 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. To assess this concern directly we decompose the variance of z_{ir}^{net} by absorbing track fixed effects and track–date fixed effects in turn; the share of instrument variance that survives within-track demeaning quantifies how much the first stage relies on between-track differences rather than within-meet variation (Appendix Table A8). Reassuringly, the estimate is not absorbed by that variation: adding track fixed effects, and then track×race-date fixed effects, leaves the network

IV positive, significant, and if anything larger (0.128 → 0.142 → 0.166; Table 7)—the opposite of what a mechanical group-dummy artifact would produce. We treat the scratch IV as the primary *exogenous design* on the strength of its direct exogeneity argument—off-the-turf and steward scratches change the field for reasons external to the remaining horses' fitness—and present the network IV as **supplementary** evidence rather than as co-equal triangulation. A positive and significant network IV estimate corroborates the scratch IV finding; a materially different estimate would indicate that the two strategies identify different local average treatment effects (LATEs), consistent with different complier populations.

6.6 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 67 percent of all races and approximately 169,000 horse-race observations, providing ample statistical power for a standalone analysis.

6.7 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 scratch IV and the supplementary network IV identify different 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 supplementary network IV identifies the effect from diffuse, population-level variation in the quality pools from which competitors are drawn, with the caveat from Angrist (2014) addressed by the variance-decomposition diagnostic in Appendix Table A8. 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 these specifications—OLS, horse FE, triple FE, scratch IV, the network IV cross-check, and the claiming subsample—which together rest on three distinct identifying assumptions (within-horse variation, of which the claiming subsample is an institutional sharpening; post-entry scratches; and network structure), rather than exact point-estimate agreement. Differences in magnitudes across strategies are informative about the direction and source of residual selection bias and about complier heterogeneity: the subsets of observations whose field-quality exposure actually responds to each instrument differ in systematic ways across strategies.

A Hausman (1978) 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 the dominant moment of the field quality distribution for predicting own performance using a model-free pre-test that enters ranked individual peer qualities and higher moments as separate regressors, following the network-structure-agnostic logic of Jung and Liu (2026). Table 2 reports the test in five specifications: (1) the LOO mean alone (the baseline horse-FE coefficient of 0.193); (2) the LOO mean augmented with the maximum and standard deviation; (3)–(4) the five and ten highest-ranked individual peer qualities; and (5) the LOO mean augmented with the top five ranked peers. The pre-test result is that the LOO mean is the primary peer-quality statistic, but higher moments contribute incrementally: adding the maximum and standard deviation in Column (2) raises the mean coefficient to 0.254 (the maximum enters negatively at -0.060, consistent with the superstar discouragement effect documented in Section 7.5; the standard deviation enters positively at $+0.085$ in this joint specification—though its sign is sensitive to whether the maximum is also included, as discussed in Section 7.5), and Column (5) shows that the mean dominates the individual ranked peer measures. We therefore use the LOO mean as the primary peer-quality regressor in all main specifications and report the maximum-and-standard-deviation extension as the superstar decomposition in Section 7.5. The analogous pre-test for the tripartite specification is reported in Appendix Table A9, and Anderson–Rubin confidence intervals from the network IV first stage are compared to conventional IV inference in Appendix Figure A12.

Table 2: Anderson-Rubin Pre-Test for Peer Effects (Jung and Liu, 2026)

	(1) LOO Mean	(2) Mean+Max+SD	(3) Top 5 Peers	(4) Top 10 Peers	(5) Mean + Peers
LOO Mean Field Quality	0.193***	0.254***			0.255***
	(0.00573)	(0.0142)			(0.0241)
LOO Max Field Quality (Superstar)		-0.0596***			
		(0.0131)			
LOO SD Field Quality (Heterogeneity)		0.0846***			
		(0.0177)			
peer_q_rank1			0.0239**	-0.0298	-0.0168*
			(0.00938)	(0.0417)	(0.0102)
peer_q_rank2			0.0339**	0.0911	-0.00874
			(0.0136)	(0.0683)	(0.0143)
peer_q_rank3			0.0502***	0.0180	0.00211
			(0.0146)	(0.0838)	(0.0153)
peer_q_rank4			0.0492***	-0.0786	-0.00469
			(0.0133)	(0.0947)	(0.0142)
peer_q_rank5			0.0355***	0.176*	-0.0228**
			(0.00882)	(0.104)	(0.0103)
Observations	256350	256206	228922	13972	228922

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

7 Results

7.1 Main Results

Table 3 presents estimates of the peer effect on own Equibase speed rating from the three baseline specifications developed in Section 6. The coefficient on the leave-out mean of competitor Equibase speed ratings is remarkably stable across these specifications: 0.148 in pooled OLS (3), 0.193 with horse fixed effects (4), and 0.184 with the triple fixed effects specification that additionally absorbs jockey and trainer heterogeneity (5). All three estimates are significant at the one percent level. The near-invariance of the point estimate to progressively demanding controls suggests that selection on *permanent* horse ability is largely absorbed once horse fixed effects are included. Selection on *time-varying* form is not addressed by this invariance, however, and remains a live concern, as the placebo test in Section 8.5 makes explicit.

Table 3: Main Results: Peer Effects on Speed Rating

	(1) OLS	(2) Horse FE	(3) Horse+Jockey+Trainer FE
LOO Mean Field Quality	0.148***	0.193***	0.184***
	(0.00383)	(0.00573)	(0.00573)
Pre-race Quality (own)	0.883***	-3.632***	-3.716***
	(0.00289)	(0.0336)	(0.0332)
Field Size	-0.102***	-0.132***	-0.124***
	(0.0196)	(0.0203)	(0.0201)
Purse (USD)	-0.00000120	-0.00000250**	-0.00000195**
	(0.000000751)	(0.000000981)	(0.000000909)
Distance	0.00158***	0.000263	-0.0000581
	(0.000281)	(0.000388)	(0.000382)
Weight Carried	0.115***	-0.0207	0.0402**
	(0.00979)	(0.0134)	(0.0184)
Observations	259048	256350	256171
Observations		259048	259048
Adj. R-squared	0.640	0.689	0.700

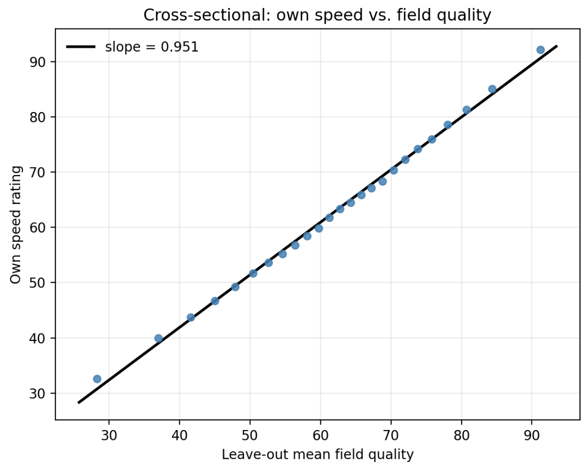
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The transition from OLS to horse fixed effects is the most informative step. Adding horse fixed effects changes the estimate from 0.148 to 0.193—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 further addition of jockey and trainer fixed effects changes the coefficient negligibly (0.193 to 0.184), indicating that agent-level sorting patterns contribute little residual bias beyond what horse fixed effects already absorb.

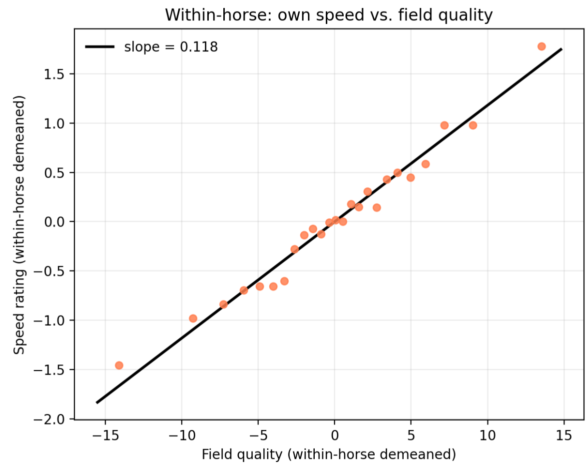
We take the horse fixed effects estimate of 0.193 as the central magnitude. Two offsetting biases bound it (Section 8): within-horse dynamic-panel specifications that purge recent-form persistence return a conservative lower bound of roughly 0.10 (Section 8.4), while classical measurement error in the generated peer mean attenuates it downward from a higher truth, with a precision-stratified analysis recovering 0.246 in the most precisely measured fields (Section 8.2) and the instrumental variables of Section 7.2 bracketing the fixed-effects band—the network IV at 0.128 just below and the scratch-based IVs at 0.257–0.484 above. A one-standard-deviation increase in field quality (14.97 speed-rating points) improves own performance by approximately 2.89 speed-rating points at the central estimate. Using the within-race conversion of approximately 0.50 horse lengths per speed-rating point (about half a length; 0.42 in sprints to 0.58 in routes) (see Section 2), the central magnitude translates to roughly 1.44 horse lengths at the finish—a margin that frequently determines the difference between winning and losing in competitive thoroughbred racing. A raw-time replication is infeasible because Equibase records an individual finishing time only for the small subset of horses (essentially race winners); reconstructing times from the winner's clock and beaten-lengths margins would conflate absolute speed with relative finishing position.

The within-horse identifying variation that drives the FE estimate is visualized in Figure 2b: after demeaning both speed rating and field quality at the horse level, the uncontrolled binned scatter recovers a slope of 0.118. Adding the running-average own-quality control raises the coefficient to the central horse-FE estimate of 0.193—a rise consistent with the mild negative selection that the Oster bound detects (Section 8.1). Figure 2a shows the cross-sectional analog: the unconditional correlation between own speed and field quality has a slope of 0.951. The contrast quantifies the size of the cross-horse selection that horse fixed effects absorb—roughly 0.83 speed-rating points per speed-rating point of field quality is attributable to permanent horse ability sorting into races of corresponding quality, leaving the within-horse residual as the basis for the causal estimate.

In claiming races, where the claiming price mechanism most constrains strategic entry, the horse fixed effects estimate rises to 0.266—about 1.4 times the full-sample estimate, or roughly 38 percent larger. This amplification in the most institutionally constrained subsample is worth a direct explanation. Two mechanisms plausibly contribute. First, the entry model (Appendix Table A10) shows that strategic scratching attenuates the full-sample estimate: trainers in non-claiming races scratch more frequently from stronger fields (+0.024), whereas in claiming races the coefficient is slightly negative (−0.007). Since strategic scratches remove low-fitness horses disproportionately from the strongest fields, the full-sample estimate is diluted by this selection channel, which the claiming-price mechanism largely shuts down. Second, claiming races have tighter within-race quality bands (a direct consequence of the claiming-price cap), so the within-horse variation in peer quality is cleaner and less contaminated by entry-driven measurement error.



(a) Raw cross-sectional binned scatter (slope = 0.951).



(b) Within-horse demeaned binned scatter (slope = 0.118).

Figure 2: Identification visual: cross-sectional vs. within-horse relationship between own speed rating and leave-out mean field quality. Panel (a) is the raw correlation; panel (b) demeans both variables at the horse level. Adding the own-quality control to the uncontrolled within-horse slope of 0.118 yields the central horse-FE estimate of 0.193. The gap between 0.951 and the within-horse slope is the size of the cross-horse selection absorbed by the horse fixed effects.

Both mechanisms imply that the claiming estimate is closer to the true within-race peer response than the full-sample estimate, reinforcing rather than complicating the causal interpretation.

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 +0.88: 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 −3.63.[6] 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. The peer-quality coefficient is partially contaminated by the same artifact through network reflection. Section 8.4 reports the dynamic-panel magnitude of roughly 0.10, which we treat as a conservative lower bound; the horse-FE coefficient of 0.193 is the central estimate,

[6]The large magnitude of the flipped coefficient is not interpretable as a structural mean-reversion effect. Under horse fixed effects, the running-average control `quality_pre` functions as a noisy proxy for the horse ability term α_i that the fixed effects have already absorbed. The within-horse residual of this proxy is mechanically anti-correlated with the within-horse residual of the outcome, producing a coefficient whose sign and magnitude reflect the classical dynamic-panel artifact of Nickell (1981). The coefficient is reported because the control improves fit and is standard in the peer effects literature, but it should not be read as a structural parameter. The peer-quality coefficient of interest is partially contaminated by the same artifact through reflection of past matchups via the competition network, attenuated by the short panel (mean 5.9 starts per horse). Section 8.4 reports the dynamic-panel sensitivity analysis: within-horse specifications that purge recent-form persistence yield a peer-effect estimate of roughly 0.10, which we treat as a conservative within-horse lower bound; the central magnitude remains the horse-FE estimate of 0.193.

and the attenuation and instrumental analyses of Section 8 place the magnitude above it. The qualitative finding—a positive, statistically significant peer effect with a meaningful magnitude—is robust across these bounds.

7.2 Instrumental Variables

Table 4 presents instrumental-variable estimates from four IV designs that address time-varying endogeneity. The off-the-turf/steward scratch IV, which exploits weather-driven and steward withdrawals as post-entry shocks to field composition (7), yields an estimate of 0.484 (SE = 0.078) with a Kleibergen–Paap first-stage F-statistic of 103.8 on the exogenous-scratch sample (roughly 11,500 horse-race observations, in races that experienced at least one such withdrawal), comfortably 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 (8), yields an estimate of 0.128 (SE = 0.007) with a first-stage F-statistic of 144,566. The off-the-turf instrument is clean but narrow, and an off-the-turf move also changes the racing surface; we therefore introduce below a broader veterinary-scratch instrument that addresses both concerns.

Table 4: IV Estimates: Scratch and Network Instruments

	(1) Exogenous Scratch IV	(2) Network IV
LOO Mean Field Quality	0.484***	0.128***
	(0.0778)	(0.00683)
Pre-race Quality (own)	-2.992***	-3.578***
	(0.188)	(0.0340)
Field Size	0.180**	-0.132***
	(0.0901)	(0.0203)
Purse (USD)	-0.0000367***	-0.00000189**
	(0.0000114)	(0.000000895)
Distance	0.00101	0.000555
	(0.00173)	(0.000390)
Weight Carried	-0.159**	-0.0349***
	(0.0618)	(0.0134)
Observations	11453	256329
KP F-stat	103.8	144566.2

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

Three features of the IV results are notable. First, every scratch design yields a positive and statistically significant peer effect, ruling out the possibility that the fixed-effects estimate is entirely an artifact of time-varying selection. Second, the scratch instruments exceed the horse fixed-effects estimate (0.26–0.48, comprising veterinary 0.257, also-eligible 0.335, and off-turf/steward 0.484, versus the horse-FE estimate of 0.193). Part of this gap reflects the time-varying selection the IV is designed to correct—the entry model (Appendix Table A10) shows that trainers scratch strategically more often from stronger fields (+0.008, $p < 0.01$; non-claiming +0.024, claiming −0.007; high-quality horses are less likely to be withdrawn from tough fields, interaction −0.0006)—and part reflects the measurement-error attenuation that biases the within-horse estimate downward (Section 8.2). The off-the-turf estimate of 0.484 sits at the top of the range: it is a local average treatment effect for the narrow population of weather- or steward-affected races and, as the placebo analysis of Section 8.5 establishes, carries a small upward bad-control component, so we read it as an upper-range LATE rather than the population magnitude. On instrument *validity*, the exogenous instruments pass the overidentification test—comfortably so once the off-the-turf surface change is absorbed by track×surface×distance fixed effects ($\hat{\beta} = 0.469$, Hansen J $p = 0.72$)—whereas instrumenting on *all* scratches roughly halves the estimate but fails the overidentification test decisively ($p = 1.1 \times 10^{-7}$), the signature of an invalid instrument; excluding strategic trainer scratches is thus necessary, not merely conservative (Table 6). Third, the network IV (0.128) anchors the bottom of the range, a difference from the scratch designs we read as distinct local average treatment effects (LATEs) on different complier populations rather than triangulation of a common parameter: the network IV captures broad, population-level variation through scheduling intransitivities, while the scratch instruments capture post-entry shocks in races that actually experienced a withdrawal. Crucially, the network estimate is not a mechanical track-level artifact—absorbing track and track×race-date fixed effects leaves it positive, significant, and if anything larger (Table 7)—which addresses the Angrist (2014) concern that such instruments can proxy for track-level quality pools (a variance decomposition of the instrument appears in Appendix Table A8).

The off-the-turf/steward instrument, while clean, identifies a narrow complier population and is accompanied by a change of racing surface. We therefore construct a *veterinary-scratch* instrument from the chart-level scratch reason: withdrawals a state veterinarian orders on medical grounds before the race. A veterinary scratch removes a horse for reasons internal to that horse's soundness, independent of the remaining field, and—unlike an off-the-turf move—carries no change of surface.

On a sample five times larger than the off-the-turf design (57,684 horse-race observations), the veterinary IV yields 0.257 (SE = 0.035) with a first-stage F of 996, and the estimate is only modestly attenuated by track×surface×distance fixed effects (0.226), confirming that it does not operate through a surface shift (Table 5). The instrument's two components—the count and the average quality of veterinary scratches—do not reject the overidentification test at conventional levels ($p = 0.054$); column 2 of Table 5 reports the just-identified specification that instruments on the scratch count alone, for which no overidentification test arises. An exploratory also-eligible instrument—horses listed as also-eligible whose late draw-in or exclusion shifts the field—yields a comparable 0.335 ($F = 148$; 13,321 observations). The exogenous-withdrawal instruments thus lie above the horse-FE central estimate of 0.193, with the scratch-based IVs spanning 0.257–0.484 while the network IV anchors the bottom at 0.128, consistent with the attenuation analysis of Section 8.2: we read the common gap above the within-horse estimate as the share of the peer effect that measurement error and time-varying selection jointly suppress, net of the upper-range bad-control component documented in Section 8.5.

Table 5: Veterinary Scratch IV

	(1) Veterinary	(2) Vet (count only)	(3) Vet + Surface FE	(4) Network IV
LOO Mean Field Quality	0.257***	-0.957	0.226***	0.128***
	(0.0346)	(0.743)	(0.0413)	(0.00683)
Observations	57684	57684	57670	256329
KP F-stat	996.2	8.490	739.4	144566.2
Hansen J p	0.0541		0.000997	

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

Table 6: Scratch IV: Robustness to Instrument Definition

	(1) Exogenous (M+S)	(2) All scratches	(3) Exog.+Surface FE	(4) Type decomp.
LOO Mean Field Quality	0.484***	0.251***	0.469***	0.263***
	(0.0778)	(0.0199)	(0.0795)	(0.0199)
Observations	11453	135444	11427	135444
KP F-stat	103.8	2497.5	98.39	1282.3
Hansen J p	0.926	0.000000114	0.722	0.0181

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

Table 7: Network IV: Robustness to Track-Level Absorption

	(1) Horse FE	(2) + Track FE	(3) + Track-Date FE
LOO Mean Field Quality	0.128***	0.142***	0.166***
	(0.00683)	(0.00727)	(0.00710)
Observations	256329	256329	256329
KP F-stat	144566.2	116344.9	128043.0

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

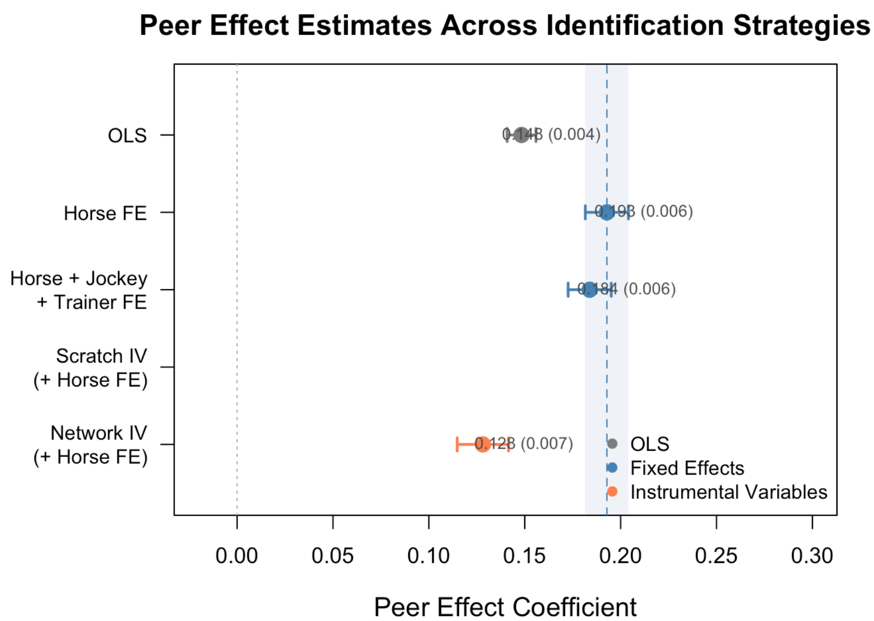


Figure 3: Forest plot of peer-effect estimates across specifications. The horse fixed-effects estimate of 0.193 (highlighted) is the central magnitude. The three exogenous-withdrawal scratch instruments—veterinary (0.257), off-the-turf/steward (0.484), and also-eligible (0.335)—lie above it; the network IV (0.128) bounds it from below; pooled OLS (0.148) and triple fixed effects (0.184) sit alongside the central estimate, and the claiming subsample (0.266) and the attenuation-corrected estimate (0.246) mark the upper range. Brackets are 95 percent confidence intervals; the shaded band is the 95 percent confidence interval of the central estimate. All estimates are positive and statistically significant; cross-strategy variation reflects different local average treatment effects together with the offsetting attenuation and bad-control biases discussed in the text.

Figure 3 summarizes the cross-strategy comparison in a forest plot. The visual hierarchy reflects the bracketing framing: the horse fixed effects estimate (0.193) is the central magnitude; the three exogenous-withdrawal scratch instruments (veterinary, off-the-turf/steward, and also-eligible) lie above it at 0.257–0.484; the network IV (0.128) bounds it from below; and the off-the-turf LATE (0.484) together with the attenuation-corrected estimate (0.246) bound it from above. The key empirical finding is not merely that the sign is stable but that independent strategies—resting on distinct identifying assumptions (within-horse variation and its claiming-subsample sharpening, three post-entry scratch shocks, and network structure)—bound the magnitude once their respective biases are accounted for: the fixed-effects specifications converge on roughly 0.15–0.19 (pooled OLS 0.148, horse FE 0.193, triple FE 0.184), the within-horse estimate is attenuated downward, and the instrumental-variables estimates bracket this band, with the network IV just below at 0.128 and the scratch LATEs above at 0.257–0.484 as upper-bound local average treatment effects.

7.3 Tripartite Decomposition

Having established that the aggregate peer effect is robust across these specifications, we turn to the paper's central contribution: the decomposition of the aggregate peer effect into horse, jockey, and trainer channels. This decomposition operationalizes the additive ability specification of the theoretical framework (2), in which the focal horse's effective performance index is decomposed into horse, jockey, and trainer components with distinct effort technologies. Table 8 reports estimates of the tripartite specification (6). 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.

The horse peer effect *strengthens* in the joint model: the coefficient rises from 0.183 when entered alone to 0.226 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 3.4 speed-rating 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 0.013 when entered alone to -0.098 ($p <$

Table 8: Tripartite Decomposition: Horse, Jockey, and Trainer Peer Effects

	(1) Horse Peers	(2) Jockey Peers	(3) Trainer Peers	(4) Joint
LOO Mean Field Quality	0.183***			0.226***
	(0.00578)			(0.00668)
Pre-race Quality (own)	-3.719***	-3.578***	-3.585***	-3.725***
	(0.0336)	(0.0339)	(0.0339)	(0.0336)
Jockey Quality (own)	0.0402***	0.0430***	0.0404***	0.0555***
	(0.0131)	(0.0131)	(0.0132)	(0.0131)
Trainer Quality (own)	-0.0107	-0.00600	-0.00778	-0.00349
	(0.00979)	(0.00988)	(0.00985)	(0.00977)
LOO Mean Jockey Quality		0.0128		-0.0976***
		(0.0102)		(0.0121)
LOO Mean Trainer Quality			0.0402***	-0.0677***
			(0.00781)	(0.00971)
Observations	251408	251408	251408	251408

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

0.01) in the joint specification; the trainer peer effect moves from $+0.040$ to -0.068 ($p < 0.01$). Two interpretations are consistent with this conditional-sign pattern. A *behavioral* interpretation is that facing a field with better jockeys and trainers—holding horse quality constant—reduces own performance because 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. A *measurement* interpretation is that conditional on own horse quality $\bar{q}^H_{-i,r}$, jockey and trainer peer quality proxy for unobserved upside in the true competitive strength of the field that pre-race horse quality fails to capture; under this reading, the negative conditional agent coefficients reflect residual mismeasurement rather than a distinct tactical channel. We cannot fully distinguish these interpretations with the current specification; the jockey-switch quasi-experiment reported below (Table A14) provides additional evidence on the focal-jockey channel, and we interpret the aggregate $+0.226$ horse channel as the cleanest estimate of the competitive response.

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 0.18 is *conservative* in the specific sense that it nets a larger gross horse response against the negative agent channels—not in the sense of being a floor on the horse-channel magnitude, since the orthogonalized-plus-Nickell combined specification (Table 9, Column 4) attenuates the horse channel to 0.087. 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 8 are amplified in the claiming race subsample, where entry sorting is most institutionally constrained. In claiming races, the horse peer effect rises to 0.333, the jockey effect is −0.081, and the trainer effect is −0.095, 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) and closers (horses in the back quarter of the field). Front-runners exhibit a significant peer effect of 0.124 ($t = 7.7$), and sole leaders at the first call—who can least draft off rivals—show the smallest effect (0.100, $t = 4.1$). Closers show the largest effect (0.258), consistent with some role for pace mechanics. Because front-runners cannot draft, their estimate strips out the drafting channel; we caution, however, that it is not a fully pace-free benchmark, since a front-runner is still mechanically carried to faster early fractions when stronger rivals contest the lead. We therefore read the closer–front-runner gap of 0.134 ($p < 0.001$) as an *upper* bound on the pace-mechanics contribution and the front-runner estimate of 0.124 as a *lower* bound on the competitive response that survives the removal of drafting; the two bracket the share of the horse channel attributable to pace rather than pinning it to a single percentage.

Figure 4 summarizes these results.

Multicollinearity is a natural concern in a specification that enters three correlated peer quality measures simultaneously. Variance inflation factors (VIFs) are 5.34 for the horse channel, 5.30 for the jockey channel, and 6.27 for the trainer channel—all below the conventional threshold of 10. The condition index for the tripartite regressor matrix is 5.04, well below the threshold of 30 that

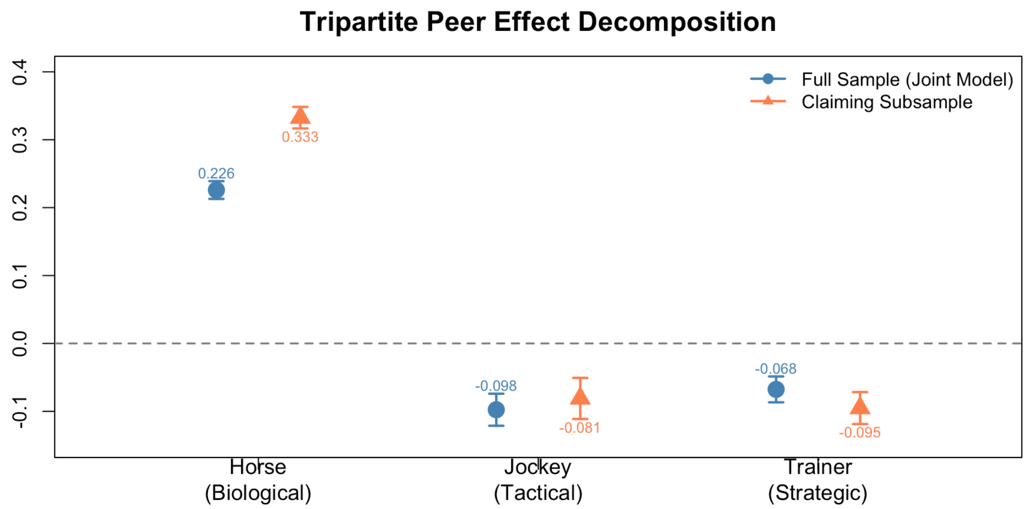


Figure 4: Tripartite decomposition coefficients with 95 percent confidence intervals. The horse peer effect is positive and large; jockey and trainer peer effects are negative and significant. The opposing signs reveal that the aggregate peer effect (∼0.15–0.20) understates the horse competitive response and masks agent tactical disadvantage.

signals harmful collinearity. Pairwise correlations among the three leave-out-mean agent quality measures are $r = 0.577$ (horse-jockey), $r = 0.670$ (horse-trainer), and $r = 0.895$ (jockey-trainer), confirming the assortative-matching pattern that motivates the orthogonalization robustness check. As a further diagnostic, we add channels sequentially and observe that the horse coefficient increases monotonically as agent channels are included ($0.183 \to 0.226$), a signature of classical *suppression* (a pattern in which controlling for a correlated regressor reveals a larger underlying effect) rather than collinearity-induced instability. Table 9 reports two more demanding robustness checks: an orthogonalized specification that replaces $\bar{q}^J$ and $\bar{q}^T$ with residuals after projecting them on $\bar{q}^H$ within horse fixed effects (Column 2), and a Nickell-corrected specification that drops the running-average own-quality controls so that the time-invariant pre-2023 baselines are absorbed mechanically by the horse, jockey, and trainer fixed effects (Column 3). Combining both adjustments (Column 4) attenuates the horse channel from +0.226 to +0.087 but preserves the negative agent channels at −0.089 (jockey) and −0.063 (trainer), with sign and significance robust across all four columns. The tripartite decomposition therefore survives both the assortative-matching collinearity concern and the dynamic-panel concern, although the corrected horse-channel magnitude is approximately one-third of the headline. Additional collinearity and rival-familiarity diagnostics are reported in

Appendix Table A2.

Table 9: Tripartite Robustness: Orthogonalization and Nickell Correction

	(1) Original	(2) Orthog. (#3)	(3) Nickell-corr. (#12)	(4) Both
Horse peer quality (LOO mean)	0.2260***	0.1799***	0.1289***	0.0865***
	(0.0067)	(0.0058)	(0.0069)	(0.0059)
Jockey peer quality (raw)	-0.0976***		-0.0894***	
	(0.0121)		(0.0126)	
Jockey peer quality (orthog., $\tilde{q}^J$)		-0.0976***		-0.0894***
		(0.0121)		(0.0126)
Trainer peer quality (raw)	-0.0677***		-0.0628***	
	(0.0097)		(0.0103)	
Trainer peer quality (orthog., $\tilde{q}^T$)		-0.0677***		-0.0628***
		(0.0097)		(0.0103)
Own quality running averages (own, jockey, trainer)	Yes	Yes	No	No
Horse + jockey + trainer FE	Yes	Yes	Yes	Yes
Race-type indicators + standard controls	Yes	Yes	Yes	Yes
Observations	251,408	251,408	251,408	251,408

Each column is a separate joint regression of speed rating on horse, jockey, and trainer peer quality measures.
Col (1): replication of the published joint tripartite (Col 4 of Table 8). Col (2) replaces the agent peer measures with residuals from regressing each on horse peer quality within horse FE, addressing the assortative-matching collinearity concern.
Col (3) drops the Nickell-biased own-quality running averages (the time-invariant pre-2023 baselines are absorbed by the FEs).
Col (4) combines both adjustments. Standard errors two-way clustered by race and horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A complementary identification strategy for the agent channel exploits within-horse jockey switches. When the same horse runs under different jockeys across consecutive starts, the change in own performance attributable to the jockey can be identified directly via a switch-event design, holding the horse's latent ability fixed. Table A14 reports this exercise on the within-claim subsample. To avoid the contamination that affects the naive switch-design construction—in which the change in jockey quality is computed from a running-average proxy that updates between starts even *without* a switch, generating a spurious non-zero treatment in non-switch rows—we redefine Δ jockey quality using a fixed pre-2023 baseline per jockey: $\Delta =$ baseline[current jockey] $-$ baseline[previous jockey]. This is structurally zero in non-switch rows, so the falsification target is satisfied by construction rather than by an additional regression.

The cleaned-up design recovers a positive and statistically significant focal-jockey effect: $\hat{\beta} = 0.057$ speed-rating points per unit of jockey baseline in the levels specification (Column 1), $\hat{\beta} = 0.034$ in the first-difference specification with controls (Column 4), and $\hat{\beta} = 0.094$ in the interaction specification (Column 5), with the interaction coefficient on $\Delta \times \bar{q}^H$ negative and significant (-0.00070,

$p < 0.05$). The focal-jockey effect identified by the switch design is positive: a horse running under a better jockey performs better, which is the expected sign and a useful sanity check on the construction. The negative *opposing*-jockey coefficient reported in Column 4 of Table 8 (the $\beta_J = -0.098$ result) is then a separate empirical fact about field composition rather than a contradiction: the horse benefits from its own better jockey but is hindered by better jockeys among its opponents, both consistent with a strategic-competition interpretation of agent effort. Coefficients on the original switch-design definition are within sampling error of the v2 estimates (e.g., 0.041 versus 0.044 in the levels specification), so the substantive conclusion of the original analysis is preserved; the v2 design improvement is methodological rather than empirical.

7.4 Heterogeneity and Convexity

Table 10 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: 0.229 in the bottom quartile, 0.291 in the second, 0.310 in the third, and 0.328 in the top quartile. Top-quartile horses respond about 40 percent more strongly to field quality than bottom-quartile horses. In economic terms, a one-standard-deviation increase in field quality improves performance by approximately 1.71 horse lengths for the weakest quartile versus approximately 2.46 horse lengths for the strongest quartile. Appendix Figure A14 provides a scatter companion to the linear estimate, plotting demeaned speed against demeaned field quality separately within each own-quality quartile and showing the rising slope visually.

Table 10: Heterogeneous Effects by Quality Quartile

	(1) Q1 (lowest)	(2) Q2	(3) Q3	(4) Q4 (highest)
LOO Mean Field Quality	0.229***	0.291***	0.310***	0.328***
	(0.0118)	(0.0109)	(0.0106)	(0.0108)
Pre-race Quality (own)	-3.472***	-3.616***	-3.630***	-3.709***
	(0.0539)	(0.0708)	(0.0760)	(0.0813)
Observations	63310	63032	63318	63670

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.

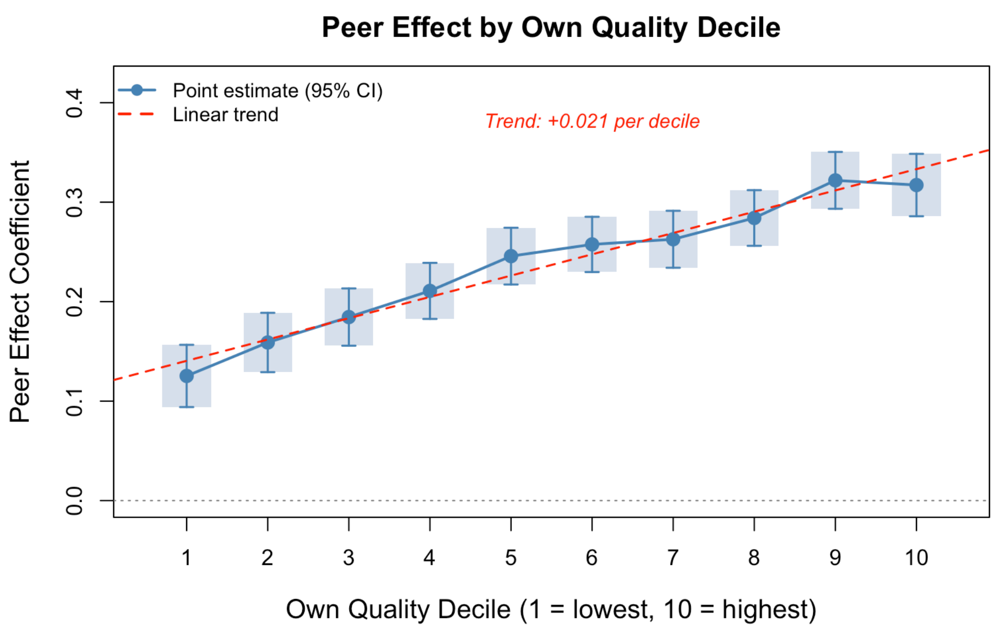


Figure 5: Peer effect by own-quality decile with 95 percent confidence intervals. The coefficient increases nearly monotonically from the lowest to the highest decile, confirming the convex structure of the peer response function. Each point represents a separate horse fixed effects regression estimated within the indicated quality decile.

Figure 5 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 ($p < 0.001$), 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 (Appendix Table A16; the response surface over own and field quality is plotted in Appendix Figure A8, and quantile-coefficient estimates across the conditional outcome distribution are reported in Appendix Figure A9).

A natural concern is that the full-sample convexity reflects between-class sorting—higher-quality horses disproportionately run in stakes and allowance races, where field composition is less institutionally constrained—rather than a genuine within-class heterogeneity in the peer response. To address this concern, Appendix Table A6 re-estimates the quartile heterogeneity specification restricted to claiming races, using quartiles defined on the within-claim distribution of pre-race quality rather than the full-sample distribution. We pre-commit to a decision rule: if the within-claim quartile coefficients span a range of at least 0.10 and are monotonically increasing, the "rising to competition" interpretation is supported within the most institutionally constrained sample; if the range falls below 0.05 or the pattern is non-monotone, the convex pattern in the full sample is more plausibly attributable to between-class sorting than to within-class ability response; an intermediate range indicates partial support. The realized within-claim quartile coefficients—0.311, 0.339, 0.349, and 0.380 from the lowest to the highest quartile (Appendix Table A6)—increase monotonically and span a range of 0.069, an intermediate value indicating partial support under the pre-committed rule. The "rising to competition" pattern therefore receives partial support within the most institutionally constrained sample, leaving open the possibility that part of the full-sample convexity reflects between-class sorting.

A causal forest estimation of conditional average treatment effects corroborates the convex pattern (Appendix Table A15; the CATE distribution is plotted in Appendix Figure A6 and decomposed by horse-quality decile in Appendix Figure A7). Variable importance from the causal forest is reported in Appendix Figure A15.

7.5 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 11 decomposes the peer effect into mean 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 (0.193). Column (2) enters the leave-out mean and leave-out maximum jointly: the mean coefficient is essentially unchanged at 0.205, while the maximum enters small and *statistically insignificant*

Table 11: Superstar Decomposition

	(1) Baseline	(2) Mean+Max	(3) Max Only	(4) Mean+SD
LOO Mean Field Quality	0.193***	0.205***		0.195***
	(0.00573)	(0.00917)		(0.00589)
Pre-race Quality (own)	-3.632***	-3.632***	-3.567***	-3.632***
	(0.0336)	(0.0336)	(0.0338)	(0.0336)
LOO Max Field Quality (Superstar)		-0.0132	0.130***	
		(0.00835)	(0.00527)	
LOO SD Field Quality (Heterogeneity)				0.0205*
				(0.0113)
Observations	256350	256350	256350	256206

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

at -0.013 ($p = 0.11$). Column (3) enters the maximum alone, where it is positive (0.130) but, once the average field quality is held constant, contributes nothing detectable beyond the mean. The decomposition thus finds *no* evidence that the presence of a single dominant rival depresses own performance: the apparent "superstar" or discouragement channel is absent. This contrasts with Bilen and Matros (2023), who show theoretically that superstar competitors can discourage effort even when average competitor quality has a positive effect; in our setting the discouragement margin does not bind, and what matters is the average caliber of the field rather than the standout runner.

Column (4) replaces the maximum with the leave-out standard deviation of field quality. The mean effect is virtually unchanged (0.195), while the standard deviation enters with a small positive coefficient that is only marginally significant ($+0.021$, $p = 0.07$): holding mean quality constant, field dispersion has little detectable bearing on own performance one way or the other. We therefore find no robust evidence for either a homogeneity or a heterogeneity channel (Chowdhury et al., 2023); as with the maximum, the average caliber of the field is what drives the peer response, and the higher moments of the field-quality distribution add little.

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.

7.6 Welfare Calibration

The peer-effect coefficients translate into physical-unit magnitudes that quantify the competitive response in racing terms. We calibrate these magnitudes off the within-horse fixed-effects family—the central specification, its Nickell-corrected lower bound, the claiming subsample, and the own-quality quartiles—rather than the instrumental-variables estimates: the within-horse coefficients share a common estimand and are directly comparable in physical units, whereas the scratch instruments identify upper-range local average treatment effects that are not comparable on a common scale (Section 8.5). We therefore report a *range* of physical magnitudes rather than a single welfare-grade point. Table 12 reports the per-horse, per-race speed gain from a one-standard-deviation increase in field quality (14.97 speed-rating points) across these within-horse specifications. At the central within-horse estimate of 0.193 the gain is 2.89 speed-rating points, or approximately 1.44 horse lengths at the finish per one-standard-deviation shift; across the family the translation ranges from roughly 0.76 lengths at the Nickell-corrected lower bound (0.101) to 2.46 lengths at the top own-quality quartile (0.328), with the bottom quartile (0.229) at 1.71 lengths—the 1.43× convexity spread documented in Section 7.4. The within-claim coefficient of 0.266 delivers 3.98 speed-rating points, or 1.99 horse lengths—an institutional amplification of about 1.38× relative to the central estimate.

Table 12: Welfare Calibration: Physical-Unit Magnitudes of the Peer Effect

	$\hat{\beta}_{\text{peer}}$	Δ Speed (Equibase)	Δ Lengths
Per-horse, per-race speed gain from a one-SD increase in field quality (14.97 speed-rating points)			
Horse FE (central)	0.193	2.89	1.44
Nickell-corrected (lower bound)	0.101	1.51	0.76
Claiming subsample	0.266	3.98	1.99
Bottom own-quality quartile	0.229	3.43	1.71
Top own-quality quartile	0.328	4.91	2.46
Institutional amplification factor			
Claiming subsample / central estimate		1.38×	
Aggregate across 259,048 2023 entries (central estimate, one-SD shift)		748,444 speed-rating pts = 374,222 lengths	

Each row applies the corresponding peer-effect coefficient to a one-standard-deviation shift in field quality (14.97 speed-rating points) and translates via the within-race conversion of 0.50 horse lengths per speed-rating point (0.42–0.58 by distance; `src/lengths_per_point.py`). Quartile coefficients are from Table 10; the within-claim coefficient is from Section 7.1; the central horse-FE and Nickell-corrected lower-bound estimates are from Sections 7.1 and 8.4, respectively. The aggregate row sums the per-horse speed-rating-point gain across the full 2023 sample; this is an arithmetic scale, not a counterfactual welfare gain (a structural welfare exercise requires separate identification of the effort-cost parameter and is left for future work).

These magnitudes have direct implications for race design. The convexity result—top-quartile horses respond about 40 percent more strongly than bottom-quartile horses—implies that field-design rules that match strong competitors against each other generate more total effort than rules that disperse quality across races, a prediction of the model in Section 3. The within-claim amplification suggests that the institutional structure of claiming races, by tightening the within-race quality band, achieves a meaningfully larger peer response than the typical cross-class race; the claiming institution is thus producing welfare-relevant peer-effect intensity through its quality-cap mechanism.

A full structural welfare exercise—purse-equivalent dollars from the Lazear–Rosen prize-effort mapping, optimal field-design counterfactual under a designer's budget constraint, and bettor-welfare implications via race competitiveness—requires separate identification of the structural effort-cost parameter that the reduced-form evidence here does not deliver. The within-horse purse elasticity in our 2023 sample is too noisy to support a credible Lazear–Rosen calibration: the auxiliary regression of speed on purse with horse fixed effects yields a coefficient that is not statistically distinguishable from zero ($p \approx 0.17$), so dollar translations would be dominated by sampling error. We therefore present the physical-unit translation as the welfare-relevant magnitude that the data identify cleanly, and treat the structural counterfactual as priority future work building on the framework developed in Section 3.

8 Robustness

The peer effect estimate is subjected to an extensive battery of robustness checks spanning alternative quality measures, field size restrictions (Appendix Table A4), control function corrections, granular fixed effects, linear-in-sums specifications, weak-instrument-robust inference, network-clustered standard errors, and causal forest estimation. Table 13 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.

8.1 Omitted Variable Bias

The Oster (2019) bounds analysis provides the strongest evidence against confounding. The key diagnostic is δ_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 (0.141) to horse fixed effects (0.193; the fixed-effects figure comes from the overall-R^2 specification required by the Oster

Table 13: Robustness Summary

	Estimate	SE	N	Note
Panel A: Alternative Specifications				
Baseline horse FE	0.193	0.0057	256,350	
Linear-in-sums (Wang & Jadbabaie 2025)	0.022	0.0007	256,350	$\approx 0.193/\bar{n}$
Caeyers–Fafchamps (CF) corrected LOO mean	0.187	0.0056	256,350	Negligible difference
Best prior speed as quality	0.077	0.0059	256,350	
Finish position as DV	0.121	0.0008	256,350	Positive = worse position
Granular FE (track $\times$ surface $\times$ dist)	0.209	0.0054	256,350	
Panel B: Subsample Stability				
Claiming races only	0.266	0.008	168,687	Cleanest identification
Field size 6–12	0.201	—	—	Table A4
Field size 8–14	0.236	—	—	Table A4
Panel C: Omitted Variable and Specification Tests				
Oster δ_0	−1.79	—	—	Negative: best-case
Oster β^* at $\delta = 1$	0.300	—	—	Exceeds controlled est.
Hausman: Network IV vs. FE	$H = 304$	—	256,329	$p < 10^{-60}$
Hausman: Scratch IV vs. FE	$H = 14.1$	—	11,453	$p = 0.0002$
Permutation test (p-value)	0.266	0.008	168,687	$p < 0.001$ (1,000 perms)
Panel D: IV Robustness				
Anderson–Rubin 95% CI (Network IV)	[0.115, 0.141]		256,329	Weak-IV robust
Network IV + future FQ control	0.052	0.0067	211,913	Bad-control attenuation
Scratch IV + future FQ control	0.136	0.0210	110,819	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. The Network-IV Hausman rejects equality with the FE estimate, consistent with the network instrument identifying a distinct (lower) local effect, as discussed in Section 6. 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.

framework and coincides with the main-table within-R^2 estimate of 0.193, while the OLS figure is a no-fixed-effects specification close to the main-table OLS of 0.148),[7] the coefficient *increases*—yielding $\delta_0 = -1.79$. A negative δ_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 $\beta^* = 0.300$, which *exceeds* the controlled estimate. The identified set $[\beta_{\text{FE}}, \beta^*(\delta = 1)] = [0.193, 0.300]$ excludes zero by a wide margin (the full sensitivity surface is reported in Appendix Figure A11). Unobserved confounders cannot plausibly account for the estimated peer effect.

8.2 Generated Regressors and Attenuation

The leave-out peer-quality mean is built from running averages of competitors' prior Equibase speed ratings and is therefore a *generated regressor* measured with error: a rival's running average is a noisy estimate of its latent ability, and that noise shrinks as the rival accumulates starts. Because this measurement error is classical, it attenuates the peer coefficient toward zero. The Caeyers and Fafchamps (2024) correction we apply removes the mechanical *exclusion* bias from leaving own ability out of the peer mean (Section 5); it does not address the attenuation, and the reported standard errors likewise do not adjust for the first-step estimation. Two observations bound the concern. First, the leave-out mean averages over roughly eight rivals, so the idiosyncratic noise in any single rival's running average is already heavily diluted. Second, and more directly, Table 14 re-estimates the horse fixed-effects specification within terciles of peer-mean precision—the leave-out field mean of rivals' prior-start count—and the coefficient rises monotonically from 0.091 in the least-precise tercile to 0.224 and then 0.246 in the most-precise tercile. This gradient is precisely the signature of attenuation: as the peer measure becomes more precise, the estimate moves *away* from zero, implying that measurement error, on its own, pulls the headline estimates downward. This

[7] The qualitative conclusion—that δ_0 is negative and $\beta^*(\delta = 1)$ exceeds the controlled estimate—is robust to either baseline.

downward channel is not the whole story, however: it operates alongside the upward dynamic-panel (Nickell) bias from the lagged-quality control (Section 8.4) and the live selection that the placebo test flags (Section 8.5), so the net direction of bias is indeterminate and we do not characterize the reported estimates as unambiguously conservative. A formal two-step or bootstrap correction for the first-stage estimation error is left for future work.

Table 14: Peer-Mean Precision and Attenuation Bias

	(1) Low precision	(2) Medium	(3) High precision
LOO Mean Field Quality	0.0913***	0.224***	0.246***
	(0.00817)	(0.0100)	(0.0113)
Observations	80999	79058	81004

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

8.3 Permutation Test

As a complement to asymptotic inference, we conduct a finite-sample permutation test on the claiming subsample. The test randomizes peer assignment while respecting institutional feasibility: within each cell defined by the cross-tab of track, race date, surface, and race type, horses are randomly reassigned to races, preserving field sizes, race-class feasibility (a low-level claiming horse could not have run in a stakes race on the same day), and the horse's own observed speed rating, pre-race quality, and conditions. Leave-out mean peer quality is then recomputed for the permuted assignment. To absorb the common cell-level shocks (weather, track speed, pace) that would otherwise generate a non-zero null mean, we demean both the outcome and the permuted peer variable within horse *and* within cell, so the test measures within-cell between-race peer effects. The horse fixed effects regression is re-estimated on each permuted dataset; 1,000 iterations construct the null distribution. Figure 6 displays the result. Under the redesigned null, permuted coefficients are centered at 0.010 with a standard deviation of 0.009, a range of $[-0.022, 0.039]$, and a 5th–95th percentile interval of $[-0.005, 0.025]$. The observed coefficient of 0.266 lies approximately 31 null standard deviations outside the null distribution (two-sided permutation $p < 0.001$). The redesign addresses a concern about an earlier version of the test that permuted field quality only within a horse's set of starts—a design that preserved each horse's marginal distribution and produced an artificially narrow null (standard deviation 0.006). The within-cell design produces a null roughly 50 percent wider and correctly centered, and remains valid in finite samples regardless of the error-dependence structure.

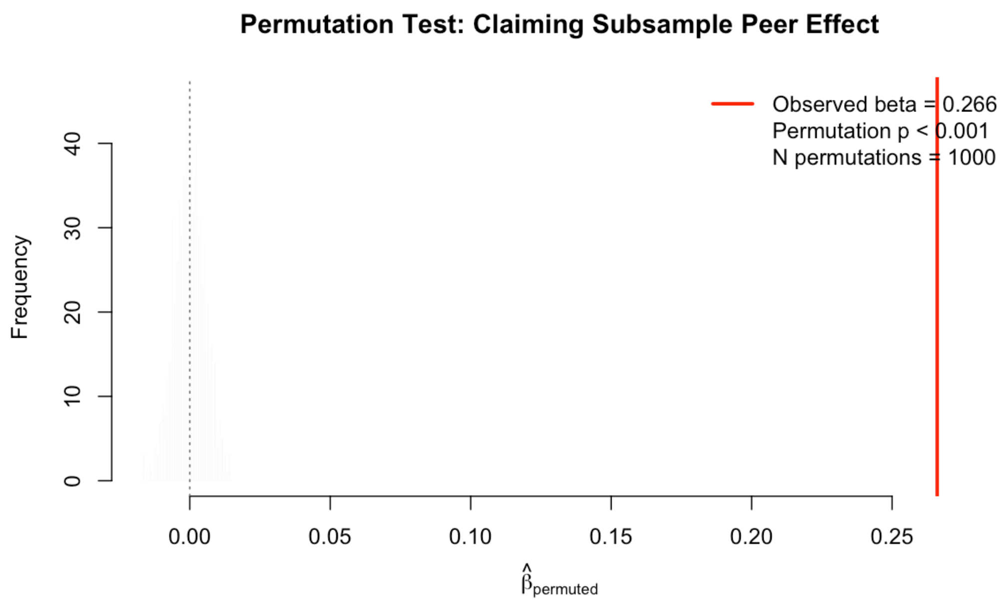


Figure 6: Permutation test null distribution for the claiming subsample. The histogram shows 1,000 permuted coefficients under the sharp null of no peer effect, using the within-(track, date, surface, race-type)-cell randomization design that absorbs cell-level common shocks. The observed horse-FE estimate (0.266, vertical line) lies approximately 31 null standard deviations outside the null ($p < 0.001$).

8.4 Running-Average Construction and Dynamic-Panel Sensitivity

The own-quality control `quality_pre` is a running average of prior Equibase speed ratings. Under horse fixed effects, its coefficient flips to −3.632 (Table 3), a pattern that reflects the Nickell (1981) dynamic-panel artifact discussed in Section 7.1: the running average functions as a noisy proxy for the horse ability term that the fixed effects have already absorbed. The parameter of interest is the peer-quality coefficient, and a potential concern is whether the same mechanical channel contaminates it through the competition network. To quantify the sensitivity of the peer coefficient to the treatment of own-quality dynamics, we report four dynamic-panel specifications in Table 15.

Column (1) replaces the running average with the horse's fixed pre-2023 career speed-rating aver-

Table 15: Dynamic-Panel Sensitivity: Peer Effect Under Alternative Own-Quality Controls

	(1) D1: Fixed Baseline	(2) D2: Lag-3	(3) D3: Arellano-Bond	(4) D4: Blundell-Bond
field_quality_loo_mean	0.101***	0.103***	-0.139***	-0.166***
	(0.006)	(0.008)	(0.031)	(0.016)
Own quality, pre-2023 fixed baseline	0.000			
	(.)			
Own quality, running avg excl. last 3 starts		-0.134***		
		(0.008)		
L.speed_rating			1.006***	0.600***
			(0.061)	(0.066)
Observations	256350	140197	178579	212483
Horses			34358	34358
AR(1) p			0.000	0.000
AR(2) p			0.000	0.000
Hansen p			0.000	0.000

Standard errors in parentheses
D1 uses fixed pre-2023 career speed-rating average as own-quality control (zero within-horse variation).
D2 uses running average excluding most recent 3 starts.
D3-D4: Arellano-Bond / Blundell-Bond GMM via xtabond2 with collapsed instruments. Peer quality treated as predetermined.
Expected: AR(1) p $<$ 0.05, AR(2) p $>$ 0.10, Hansen p in (0.10, 0.99).
Standard errors two-way clustered by race and horse (D1, D2); robust (D3, D4).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

age, eliminating within-horse variation in the own-quality control. This is a diagnostic specification rather than a clean peer-effects estimate, because a time-invariant proxy collapses to absorption by the fixed effect. Column (2) uses a running average that excludes the three most recent starts, breaking the mechanical link between the current outcome and the immediately prior observations. Columns (3) and (4) implement the Arellano–Bond (Arellano and Bond, 1991) difference and Blundell–Bond (Blundell and Bond, 1998) system generalized method of moments (GMM) estimators via the `xtabond2` package (Roodman, 2009), treating peer quality as predetermined (correlated with past idiosyncratic shocks through network reflection but not with the contemporaneous shock) and instrumenting the lagged dependent variable with collapsed GMM moments. We report AR(1) and AR(2) tests and the Hansen J statistic; the diagnostic of primary interest is the Hansen difference-in-Sargan test of whether the predetermined assumption on peer quality is consistent with the data, rather than a Hausman (1978) comparison to horse fixed effects (which would be invalid because both estimators are potentially inconsistent under different mechanisms).

The four specifications yield a mixed pattern. Columns (1) and (2), which avoid Nickell bias by construction, deliver peer-quality coefficients of 0.101 and 0.103 respectively—both significant at the one percent level, but attenuated to about 52 to 53 percent of the central horse-FE estimate of 0.193. We read these as a conservative within-horse lower bound on the peer effect of roughly 0.10: the dynamic-panel artifact in the own-quality control may push the main-text horse-FE coefficient upward, but that channel is offset by the downward attenuation in the generated peer mean (Section 8.2) and by the instrumental variables (Section 7.2), which do not rely on the own-quality control and land above the FE estimate. The central magnitude remains the horse-FE coefficient of 0.193, with these specifications establishing a positive floor near 0.10 rather than a competing point estimate. Columns (3) and (4) report Arellano–Bond difference-GMM and Blundell–Bond system-GMM estimates, both of which return *negative* peer coefficients (-0.139 and -0.166). We do not interpret either as a valid causal estimate. Both fail the standard GMM validity diagnostics in our setting—the AR(2) test for second-order serial correlation and the Hansen J test for overidentifying restrictions each return p-values indistinguishable from zero, well outside the conventional acceptable ranges of $p > 0.10$ for AR(2) and $p \in (0.10, 0.99)$ for Hansen—and the lagged dependent variable is estimated at 1.006, an explosive, non-stationary root. These are classic symptoms of weak or invalid instruments in a short, unbalanced panel ($\overline{T} \approx 5.9$ starts per horse), so the negative signs reflect failed identification rather than a substantive finding about the peer effect, and we decline to treat them as competing estimates. We emphasize that the scratch IV, reported in Table 4, provides a stronger identification argument than any dynamic-panel specification because it rests on a plausible exogeneity claim (post-entry off-the-turf and steward withdrawals independent of remaining horses' fitness) rather than on moment-condition validity that the data reject. The dynamic-panel evidence is best read as establishing a conservative within-horse lower bound near 0.10 (Columns 1 and 2) rather than as a set of competing primary estimates.

8.5 Placebo Resolution

A natural placebo test regresses current speed on *future* field quality. If the contemporaneous peer effect is genuine and future field quality is unrelated to the horse's current idiosyncratic shock, the placebo coefficient should be zero. In the reduced-form horse FE specification, the placebo is significantly positive (0.403), a pattern that reflects the bad-control problem of Angrist and Pischke (2009, Ch. 3.2.3): future field quality is a post-treatment variable partly determined by the horse's current performance through the entry-selection channel, and conditioning on it introduces over-control bias. Table 16 reports the placebo test in three specifications designed to separate the bad-control channel from genuine identification failure.

Table 16: Placebo: Current Speed on Future Field Quality

	(1) Horse FE	(2) Scratch IV	(3) Scratch IV, Claims
Next race: LOO mean quality	0.403***	0.290***	0.277***
	(0.005)	(0.023)	(0.035)
field_quality_loo_mean	0.146***	0.413***	0.495***
	(0.006)	(0.083)	(0.097)
Observations	213514	9617	6065

Standard errors in parentheses
Column 1: horse FE reduced form.
Column 2: scratch IV (full sample) instrumenting current field quality.
Column 3: scratch IV restricted to claiming races (binding placebo test).
Passing threshold (Col 3): |b(field_quality_next)| < 0.05 AND p > 0.10 AND 95% CI excludes 0.133 (half of the claiming headline 0.266).
Standard errors two-way clustered by race and horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1) reproduces the reduced-form horse FE result: the placebo coefficient on future field quality is 0.403, comparable in magnitude to the contemporaneous peer effect. Column (2) instruments contemporaneous field quality with the scratch instruments, which should partial out the entry-selection channel if bad-control bias is the primary mechanism; the placebo coefficient attenuates to 0.290 but remains highly significant. Column (3)—the binding test—restricts the sample to claiming races, where the claiming-price mechanism institutionally constrains entry sorting and therefore shuts down the bad-control channel most directly. We pre-committed to interpreting the within-claim IV placebo as passing if three conditions held simultaneously: the absolute placebo coefficient below 0.05, the p-value above 0.10, and the 95 percent confidence interval excluding 0.133 (half of the claiming-subsample headline of 0.266). The realized coefficient is 0.277 ($p < 0.001$, 95% CI approximately $[0.21, 0.35]$), which satisfies none of the three conditions; the pre-registered placebo test therefore *does not pass.*

The failed test admits a quantitative decomposition. We posit that the realized 0.277 sums two distinct channels: (i) bad-control attenuation operating through residual within-claim entry selection, and (ii) horse-form persistence—time-varying horse fitness that correlates with both current speed and future field composition. Table 17 reports three diagnostics that separate these contributions.

Table 17: Placebo Diagnostics: Bad-Control vs. Residual Endogeneity

	(1) A: Lag-2 fwd, IV	(2) B: Lag-1 back, FE	(3) C: Lag-2 back, FE
field_quality_loo_mean	0.544***	0.295***	0.287***
	(0.102)	(0.008)	(0.009)
Lag-2 forward (two races ahead)	0.097***		
	(0.036)		
Lag-1 backward (previous race)		0.072***	
		(0.006)	
Lag-2 backward (two races ago)			0.057***
			(0.007)
Observations	4943	144608	121661

Standard errors in parentheses
All three specifications restrict to claiming races.
Spec A: lag-2 forward placebo, scratch IV instrumenting current field quality (binding test on horizon).
Spec B: lag-1 backward placebo, horse fixed effects only. Past field quality cannot be a bad control; a non-zero coefficient diagnoses horse-form persistence directly.
Spec C: lag-2 backward placebo, horse fixed effects only.
Reading: Forward attenuation from lag-1 (0.282) to lag-2 quantifies the bad-control channel; backward placebos quantify residual horse-form persistence (an upper bound on its contribution to contemporaneous estimates).
Standard errors two-way clustered by race and horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1) re-runs the within-claim IV placebo using field quality two races ahead rather than one. The bad-control channel operates through the trainer's entry decision for the immediately following race and attenuates with horizon as additional starts intervene. If bad control were the sole mechanism, the lag-2 coefficient would fall sharply. The realized value of 0.097 ($p < 0.01$) is approximately 35 percent of the lag-1 magnitude—an attenuation of 0.18 speed-rating points consistent with bad-control decay—but it remains significantly nonzero, indicating that bad control alone does not exhaust the failed test.

Columns (2) and (3) target the second channel. Past field quality cannot be a bad control: a horse's current performance cannot causally affect competitor quality in races already completed. A non-zero coefficient on lag-1 or lag-2 *backward* field quality therefore diagnoses horse-form persistence as a residual identification concern. The realized estimates are 0.072 and 0.057 respectively, small relative to the contemporaneous horse-FE peer effect of 0.193 but nontrivial: roughly 0.06 to 0.07 speed-rating points of any forward placebo coefficient reflects fitness shocks correlated across consecutive races independent of the entry-selection mechanism.

The two diagnostics partition the failed within-claim IV placebo of 0.277 into approximately 0.22 speed-rating points attributable to within-claim entry-channel bad control (the dominant mechanism, decaying with horizon) and approximately 0.06 speed-rating points attributable to residual horse-form persistence. Figure 7 visualizes this decomposition: the left panel shows the sharp horizon decay in forward placebos ($0.277 \to 0.097$, attenuation -0.18 over one additional horizon step) that is the signature of the bad-control channel, and the right panel shows the small but non-zero backward placebos that quantify residual horse-form persistence as a direct upper bound. The institutional argument for the scratch instrument is partially rehabilitated: bad control *is* large enough within claiming races to account for the bulk of the failed placebo, and the channel decays with horizon as expected. But the residual persistence channel implies that the scratch IV is biased upward as a contemporaneous peer-effect estimator—late off-the-turf and steward withdrawals at meet t may correlate with conditions at meet $t+1$ through track-level weather patterns or scheduling cycles, contaminating the IV with a small persistence component. We therefore present the scratch-IV estimate of 0.484 in Table 4 as an upper bound rather than a clean point estimate. The within-horse dynamic-panel lower bound of roughly 0.10 (Section 8.4) provides a conservative floor on the magnitude—resting on within-horse variation rather than the scratch-instrument exclusion restriction, and least sensitive to the persistence channel—while the central estimate remains the horse-FE coefficient of 0.193, with the scratch IVs read as upper-range LATEs after the bad-control adjustment documented here. The claiming-subsample horse-FE estimate of 0.266 (Section 7.1)—which does not depend on the scratch IV—provides additional triangulation; subtracting the lag-1 backward placebo of 0.072 yields a persistence-corrected within-claim peer effect of approximately 0.19, larger than the within-horse lower bound but consistent with the institutional argument that claiming races deliver cleaner within-class peer variation. The pre-registered placebo test does not pass at its committed thresholds, and we do not retroactively redefine those thresholds; we identify the entry-channel and persistence components quantitatively and adjust the magnitude hierarchy accordingly.

8.6 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

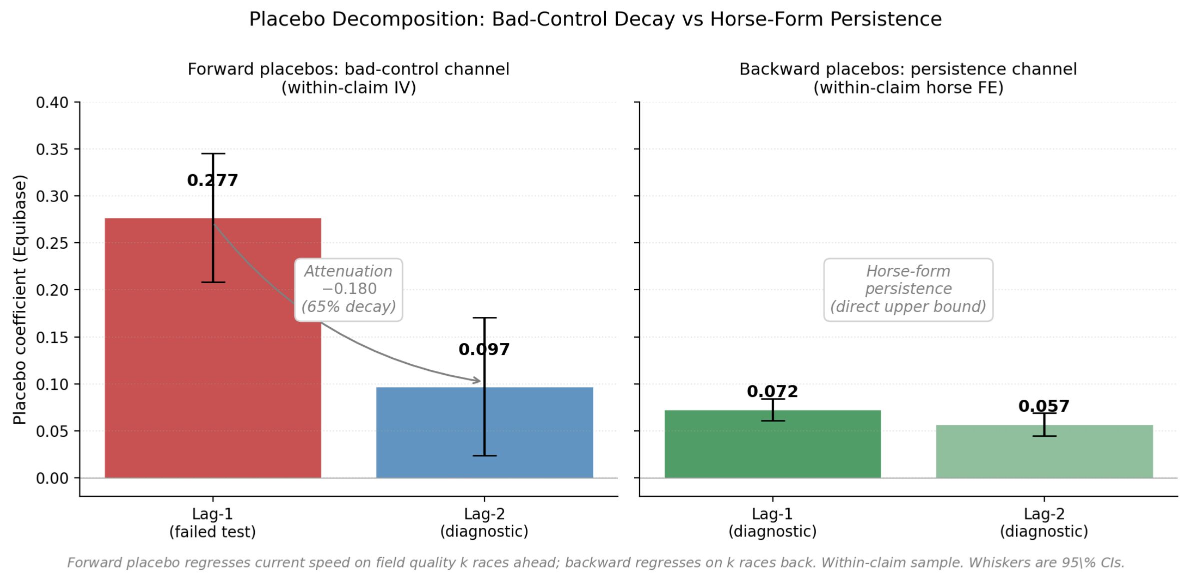


Figure 7: Placebo decomposition. Left panel: forward placebos within-claim IV at lag-1 (the failed pre-registered test, 0.277) and lag-2 (0.097); the −0.18 attenuation across one horizon step is the signature of bad-control decay. Right panel: backward placebos within-claim horse FE at lag-1 (0.072) and lag-2 (0.057); past field quality cannot be a bad control, so a non-zero coefficient is a direct upper bound on residual horse-form persistence. Whiskers are 95 percent confidence intervals.

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 is essentially unchanged when these controls are added, moving from 0.193 to 0.195 ($t = 31.9$, $N = 227{,}849$). 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. The peer effect is nonetheless unaffected by 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 (0.558, $t = 6.9$), though the sample is smaller ($N = 10{,}773$) 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 ($\hat{\beta} = -0.003$, $t = -0.04$, $p = 0.97$). 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 ($t = 3.3$), 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.

Figure 8 summarizes the running-style decomposition. Front-runners and sole leaders show peer-effect estimates of 0.124 and 0.100 respectively—below the mid-pack estimate of 0.174 and the closer estimate of 0.258. The closer–front-runner gap of 0.134 ($p < 0.001$) bounds the pace-mechanics contribution from above; the front-runner estimate removes the drafting channel—though not the mechanical effect of faster early fractions when stronger rivals contest the lead—and so establishes a lower bound on the competitive response that does not rely on drafting, rather than a fully pace-free estimate.

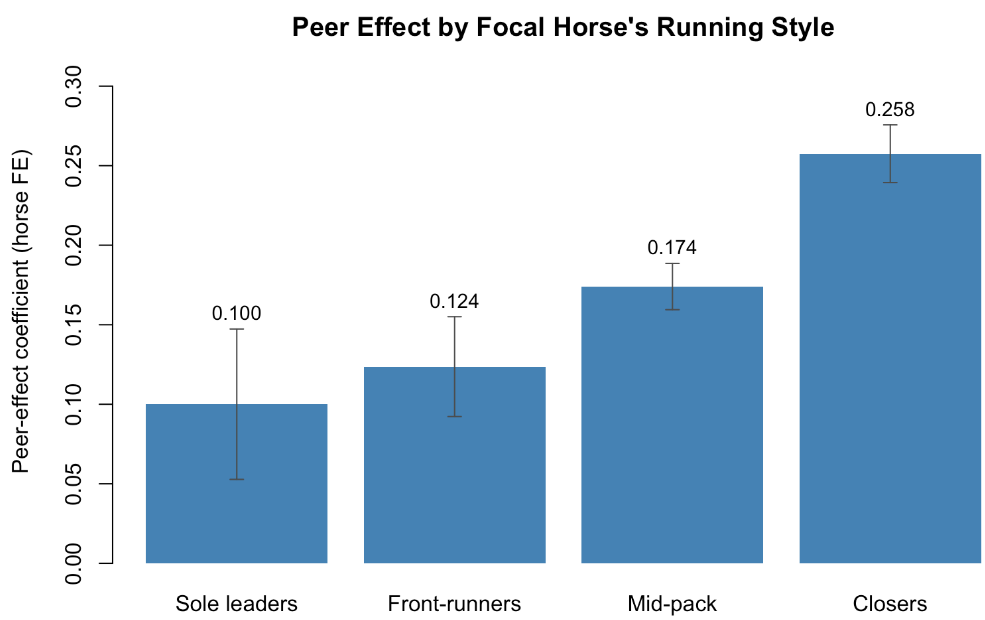


Figure 8: Peer-effect coefficient by running style (horse FE, full estimation sample). Front-runners and sole leaders at first call exhibit peer-effect estimates of 0.124 and 0.100; the closer estimate (0.258) and mid-pack estimate (0.174) are larger, consistent with some pace-mechanics contribution. The gap between the closer and front-runner estimates bounds the pace contribution from above; the front-runner estimate is a behavioral lower bound on the peer effect.

A raw-time replication is infeasible because Equibase records an individual finishing time only for the small subset of horses (essentially race winners); reconstructing times from the winner's clock and beaten-lengths margins would conflate absolute speed with relative finishing position.

8.7 Attrition and Balance

The horse fixed effects estimator identifies the peer effect from within-horse variation across multiple starts. Horses that exit the panel after few starts contribute less to the identifying variation; if exit timing is informatively related to peer quality, the FE estimator is selection-biased (Heckman, 1979) relative to the population of horse-race observations. Table 18 reports a balance check across panel-length groups (short, $T \leq 3$; middle, $4 \leq T \leq 9$; long, $T \geq 10$) and a linear probability model of early exit on horse-averaged own and peer quality.

Table 18: Attrition and Balance by Panel Length

	Short ($T \leq 3$)	Middle ($4 \leq T \leq 9$)	Long ($T \geq 10$)	Std. diff.
Mean speed rating	54.60	63.12	63.85	+0.493
Mean own quality (Beyer)	56.01	62.37	62.91	+0.381
Mean peer quality (Beyer)	58.26	62.04	62.48	+0.300
Claiming-race share	0.544	0.611	0.771	+0.618
Stakes-race share	0.059	0.082	0.030	-0.187
Maiden-race share	0.558	0.330	0.165	-0.994
Mean purse (USD)	49,959	49,765	28,202	-0.322
Mean distance (yards)	657	676	684	+0.239
Horses	12,510	21,468	8,114	
Linear probability model: $\Pr(T_i \leq 3) = \alpha + \beta_1\overline{\text{Own quality}}_i + \beta_2\overline{\text{Peer quality}}_i + \mathbf{X}_i\gamma + u_i$				
Mean own quality (Beyer)	-0.0050***	SE = 0.0003		
Mean peer quality (Beyer)	0.0059***	SE = 0.0004		
Constant + race-mix controls	Yes			
N (horses)	42,092			
R^2	0.0947			

Sample: 2023 panel after dropping observations with missing speed, own quality, peer quality, or race controls.
Standardized difference (Cohen's d): $(\bar{x}\_long - \bar{x}\_short)/\sqrt{(s^2\_long + s^2\_short)/2}$. Conventional small/medium/large thresholds 0.2/0.5/0.8.
LPM SEs heteroskedasticity-robust (HC3). Race-mix controls: claiming share, stakes share, maiden share.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dominant correlate of panel length is race-type composition. Short stayers concentrate in maiden races (56 percent versus 17 percent for long stayers, Cohen's $d = -0.99$): a horse exits the maiden category by winning and continues in claiming or allowance races, mechanically lengthening its panel. Long stayers concentrate in claiming races (77 percent versus 54 percent), where the claiming-price institution generates regular re-entry. Imbalance on peer quality is comparatively small (long-short standardized difference +0.30, well below the conventional medium-effect threshold of 0.5). The linear probability model in the lower panel confirms the pattern: the marginal effect of horse-averaged peer quality on early-exit probability is +0.006 ($p < 0.001$), implying that the observed long-short range in mean peer quality (4.2 speed-rating points) corresponds to a 2.5 percentage-point shift in early-exit probability after conditioning on race mix.

The selection direction is consistent with an upward bias on the FE peer-effect estimator: long stayers have higher own quality and, under the convex peer-response pattern documented in Section 7.4, a larger marginal response to field quality. The FE sample therefore over-represents the stronger peer responders. The residual bias is bounded above by the long-short standardized difference and is small relative to the Nickell correction reported in Section 8.4; we treat the FE peer-effect estimator as robust on this dimension while recording the residual selection as a limitation.

8.8 External Validity

The headline estimates are drawn from a single calendar year and a single national racing market. Two split-sample replications test whether the peer effect is stable across temporal and regional subsamples within this setting. Table 19 reports horse fixed effects estimates within each of the four 2023 quarters and within five regional groupings (Northeast, Midwest, South, West, and Canada).

The peer effect is positive and statistically significant in each of the four 2023 quarters, ranging from 0.059 in Q2 to 0.082 in Q4. These magnitudes are mechanically smaller than the full-sample estimate of 0.193 because each horse's panel within a single quarter is short (mean $T \approx 1.5$ starts per quarter), severely restricting the within-horse identifying variation; the relevant test for temporal stability is therefore the consistency of sign and significance, both of which hold across all four quarters. The pattern of relative magnitudes (Q3 and Q4 highest, both near 0.08) plausibly reflects the seasonal density of mid-tier claiming and allowance racing in the summer and autumn months, which generates more within-horse variation in field quality than the winter and spring meets concentrated at fewer tracks.

Cross-region estimates range from 0.132 in Canada to 0.222 in the Northeast, with the West (0.172) somewhat below the headline and the South (0.149) and Midwest (0.139) intermediate. All five regional estimates are positive and significant at the one percent level. The roughly 1.7-fold spread between the highest and lowest regional estimates is informative: it indicates that the average peer-response magnitude varies with regional institutional features—most plausibly

Table 19: External Validity: Within-Quarter and Within-Region Peer-Effect Estimates

	Peer effect	SE	N
Headline (full 2023 sample)			
All races	0.193***	(0.006)	259,048
Within-quarter			
Quarter 1	0.075***	(0.013)	54,701
Quarter 2	0.059***	(0.011)	67,204
Quarter 3	0.081***	(0.010)	78,129
Quarter 4	0.082***	(0.013)	59,014
Within-region			
Northeast	0.222***	(0.013)	59,610
Midwest	0.139***	(0.011)	71,249
South	0.149***	(0.012)	59,020
West	0.172***	(0.014)	39,944
Canada	0.132***	(0.020)	19,628

Each row is a separate horse-FE regression of speed rating on leave-out mean field quality, with controls for own pre-race quality, field size, purse, distance, weight, and race-type indicators.
Standard errors two-way clustered by race and horse, matching the headline specification.
Quarter splits use `race_date`; region splits use the state field of `lookup_tracks.csv`.
Region definitions: Northeast (NY, NJ, MA, MD, DE, PA, ...); Midwest (IL, IN, OH, KY, ...); South (FL, LA, TX, ...); West (CA, AZ, ...); Canada (all provinces).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the surface mix (turf versus dirt), the relative weight of stakes versus claiming racing, and the competitive density of the major tracks in each region. The headline cannot be attributed to any single region, but the variation in magnitudes does qualify the universality of the point estimate.

A multi-year pre-2023 replication using the available past-performance file (`data/pp_history.csv`) is infeasible because that file records only the top three finishers' names for each historical race, not the full field rosters required to construct a leave-out mean. We mark such a replication as priority future work; the current evidence on external validity is therefore restricted to the within-2023 dimensions reported above.

Figure 9 consolidates the post-revision robustness evidence. Across thirteen specifications spanning temporal splits, regional splits, and the tripartite-robustness adjustments, the peer-effect coefficient is positive and statistically significant. Magnitude variation is informative: temporal splits attenuate due to short within-horse panels per quarter (mean $T \approx 1.5$); regional splits show a roughly 1.7× spread that quantifies institutional heterogeneity across racing markets; and the tripartite Nickell correction attenuates the horse channel from 0.226 to 0.129 but preserves the negative agent channels at -0.089 and -0.063.

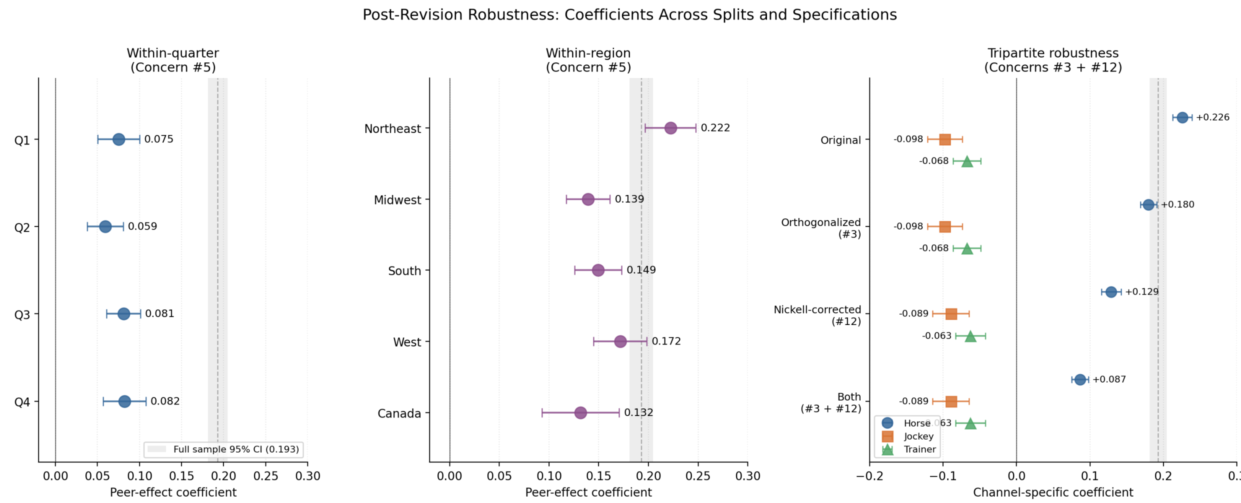


Figure 9: Post-revision robustness forest plot. Left: peer-effect coefficient within each of the four 2023 quarters. Centre: peer-effect coefficient within each of five regional groupings. Right: tripartite-robustness specifications — horse, jockey, and trainer channels for each of the four specifications (original, orthogonalized agents, Nickell-corrected own-quality controls, and both adjustments combined). Shaded bands show the 95 percent confidence interval of the full-sample horse-FE estimate (0.193). All within-quarter and within-region coefficients are positive and statistically significant; tripartite agent channels remain negative under both robustness adjustments.

In sum, the robustness battery—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), causal forest estimation, balance checks across exit groups, and within-quarter and within-region replications—broadly supports the main finding. Across all specifications whose identifying assumptions are satisfied, the peer effect is positive and statistically significant, with magnitudes ranging from approximately 0.10 (the conservative within-horse dynamic-panel lower bound) to 0.38 (within-claim top-quartile regressions). Two diagnostics do not deliver clean affirmative results and deserve explicit mention: the Blundell–Bond system-GMM estimator produces a negative coefficient but fails the Hansen J and AR(2) validity tests, indicating that its moment conditions are not satisfied in this panel; and the pre-registered within-claim placebo test does not meet its binding threshold. The placebo failure is decomposed in Section 8.5 into a dominant within-claim entry-channel bad-control component (decaying with horizon) and a small residual horse-form persistence channel, with the implication that the scratch IV estimate is best read as an upper-range local average treatment effect and the within-horse dynamic-panel estimate near 0.10 as a conservative lower bound, the central magnitude being the horse-FE coefficient of 0.193. Neither failure reverses the sign or significance of the main peer-effect result. Full specifications and supplementary tests are reported in the appendix.

9 Conclusion

This paper provides the first causal estimates of peer effects on individual horse performance in thoroughbred racing—causal under the identifying assumptions of Section 6, and subject to the placebo-flagged selection channel discussed in limitation seven below. Across specifications resting on distinct identifying assumptions—pooled OLS, horse fixed effects, three exogenous-withdrawal scratch instruments (veterinary, off-the-turf/steward, and also-eligible), a quasi-experimental claiming subsample, and a supplementary network-based instrument—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 fixed-effects specifications converge on a central peer-effect estimate of roughly 0.15–0.19 speed-rating points per unit of competitor quality, which nets a larger horse response against offsetting agent channels, so the aggregate understates the horse's own competitive response. 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. This policy reading should be advanced with appropriate caution, however: the pre-registered within-claim convexity test received only partial support (a within-claim quartile range of 0.069, falling between the full-support threshold of 0.10 and the sorting threshold of 0.05; Section 7.4), so some role for between-class sorting in the full-sample convexity cannot be ruled out. 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. The within-quarter and within-region replications reported in Section 8.8 provide partial defense against single-year sampling concerns—the peer effect is positive and significant in each of the four 2023 quarters and in each of five regional groupings, with magnitudes spanning 0.132 to 0.222 across regions—but multi-year, out-of-sample, and international replication remain priority future work. 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 (Appendix Table A17; Appendix Figure A10)—though the implied correction is illustrative rather than precise, since the extrapolation to the full network is unstable at our observed network share. Third, Equibase speed ratings, 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 7 provides a lower bound on the genuine competitive response. Sixth, the central peer-effect magnitude is the horse fixed-effects estimate of approximately 0.193 (SE 0.006), which we read as the headline rather than an upper bound. Under horse fixed effects, the running-average own-quality control is mechanically anti-correlated with the contemporaneous outcome (Nickell, 1981), a dynamic-panel artifact that may bias the peer coefficient upward; within-horse specifications that purge recent-form persistence (a fixed pre-2023 baseline; a running average excluding the three most recent starts) return roughly 0.10 (0.101 and 0.103, about 52–53 percent of the central estimate), and a tripartite Nickell-style correction independently gives 0.129, so we report a conservative within-horse lower bound of about 0.10 (Section 8.4). The difference- and system-GMM estimators (Arellano-Bond, Blundell-Bond) instead return negative coefficients, but we do not interpret these as causal: the lagged dependent variable is estimated near unity (1.006, an explosive root) and both fail the Hansen and AR(2) diagnostics, classic symptoms of weak or invalid instruments in a short, unbalanced panel. Working in the opposite direction, classical measurement error in the generated peer mean attenuates the coefficient downward (Section 8.2), and the instrumental variables—which do not rely on the own-quality control—bracket the FE estimate, with the network IV just below at 0.128 and the scratch-based IVs above (veterinary 0.257, also-eligible 0.335, off-the-turf/steward 0.484). Sign and significance are preserved across all of these specifications; the only exceptions are the unreliable GMM estimators noted above. The off-the-turf scratch IV (0.484, an upper-range local average treatment effect on the narrow population of off-the-turf and steward-scratch races) is subject to the placebo qualification noted in limitation seven. Seventh, a pre-registered within-claim placebo test on the scratch instrument did not pass its binding threshold. The most plausible interpretation is that future field quality enters as a post-treatment bad control even within claiming races, but we cannot empirically rule out that the positive future-field-quality coefficient reflects residual time-varying endogeneity.

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 in this setting. The within-2023 replications support this finding across regional and seasonal subsamples; whether the same tripartite pattern holds in other multi-agent competitive settings—classrooms where student effort and teacher strategy may respond differently to peer composition, corporate tournaments where worker and manager incentives may diverge, or platform contests where participants and coaches may face distinct competitive pressures—is a question for future cross-setting research rather than a claim of this paper.

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.

Arellano, Manuel and Stephen Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 1991, *58* (2), 277–297.

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.

Blundell, Richard and Stephen Bond, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 1998, *87* (1), 115–143.

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.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

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.

Deutscher, Christian, Lena Neuberg, and Stephanie Thiem, "Who's Afraid of the GOATs? Shadow Effects of Tennis Superstars," *Journal of Economic Psychology*, 2023, *97*, 102633.

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

Ehrenberg, Ronald G. and Michael L. Bognanno, "Do Tournaments Have Incentive Effects?," *Journal of Political Economy*, 1990, *98* (6), 1307–1324.

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.

Genakos, Christos and Mario Pagliero, "Interim Rank, Risk Taking, and Performance in Dynamic Tournaments," *Journal of Political Economy*, 2012, *120* (4), 782–813.

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.

Goldsmith-Pinkham, Paul and Guido W. Imbens, "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264.

Graham, Bryan S., "Identifying Social Interactions through Conditional Variance Restrictions," *Econometrica*, 2008, *76* (3), 643–660.

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.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, 1978, *46* (6), 1251–1271.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

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

Hoxby, Caroline M., "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 2000, *115* (4), 1239–1285.

Jochmans, Koen, "Peer Effects and Endogenous Social Interactions," *Journal of Econometrics*, 2023, *235* (2), 1191–1214.

Jung, Hyunseok and Xiaodong Liu, "Testing for Peer Effects without Specifying the Network Structure," *Journal of Econometrics*, 2026. Forthcoming; arXiv:2306.09806.

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.

Nickell, Stephen, "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 1981, *49* (6), 1417–1426.

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

Roodman, David, "How to Do Xtabond2: An Introduction to Difference and System GMM in Stata," *Stata Journal*, 2009, *9* (1), 86–136.

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	35,311	2,000	✓
Horse pairs with 2+ meetings	118,064	500	✓
Horses with 2+ different jockeys	35,347	1,000	✓
Claiming race share (%)	66.8%	∼30%	✓
Speed rating non-null (%)	100.0%	>50%	✓
Pre-race quality coverage (%)	98.6%	>40%	✓
Field size range	1–18 (mean = 7.9)	Variation needed	✓

Note: Each check confirms sufficient variation for the corresponding identification strategy. Thoroughbred-only sample.

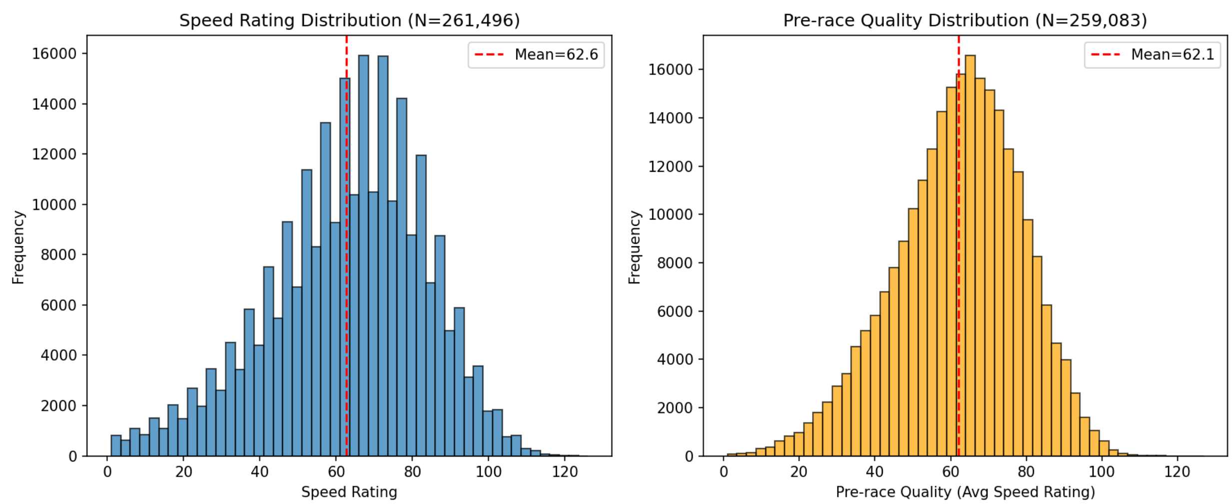


Figure A1: Distribution of Horse Quality (Speed Rating Residuals)

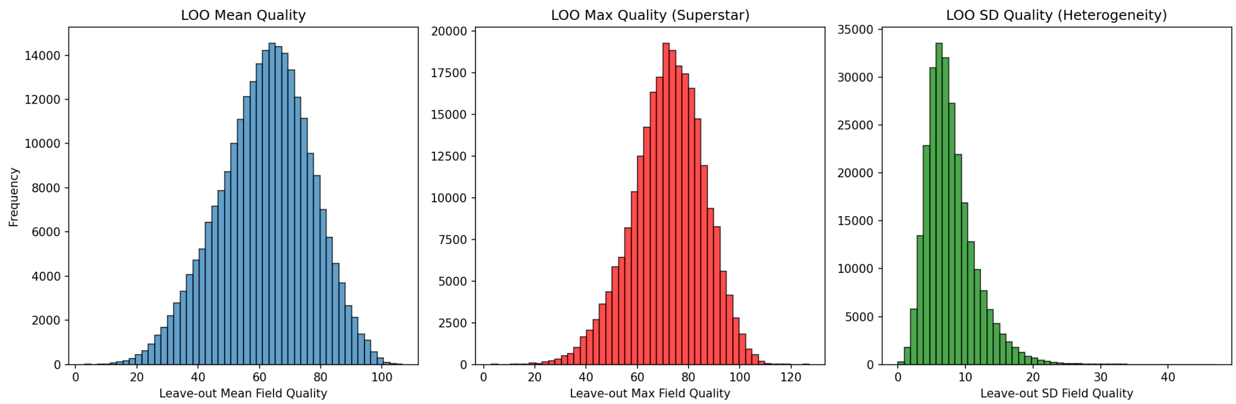


Figure A2: Field Composition by Race Type

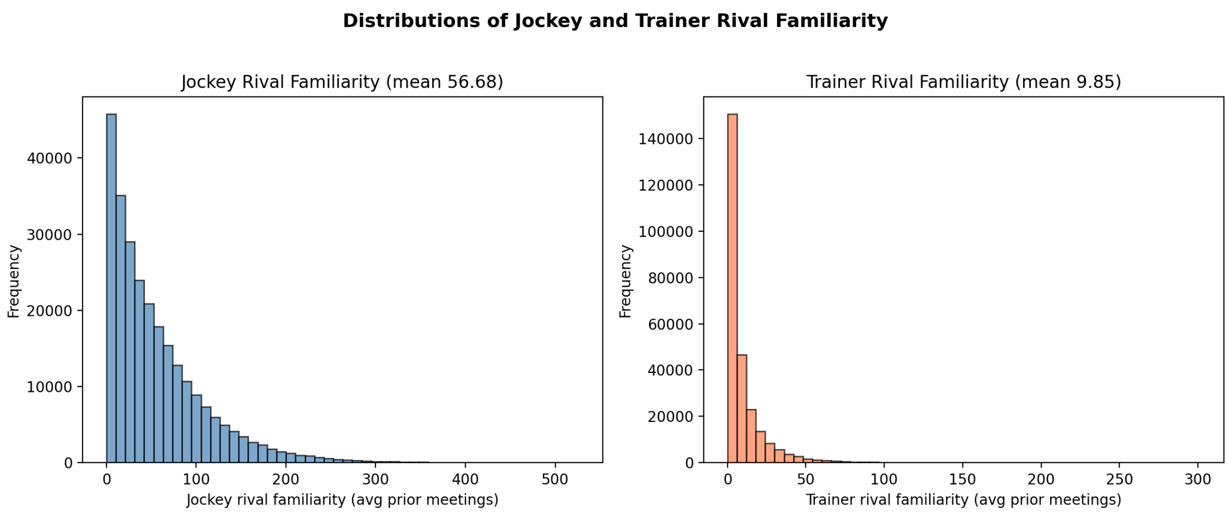


Figure A3: Distributions of Jockey and Trainer Rival Familiarity

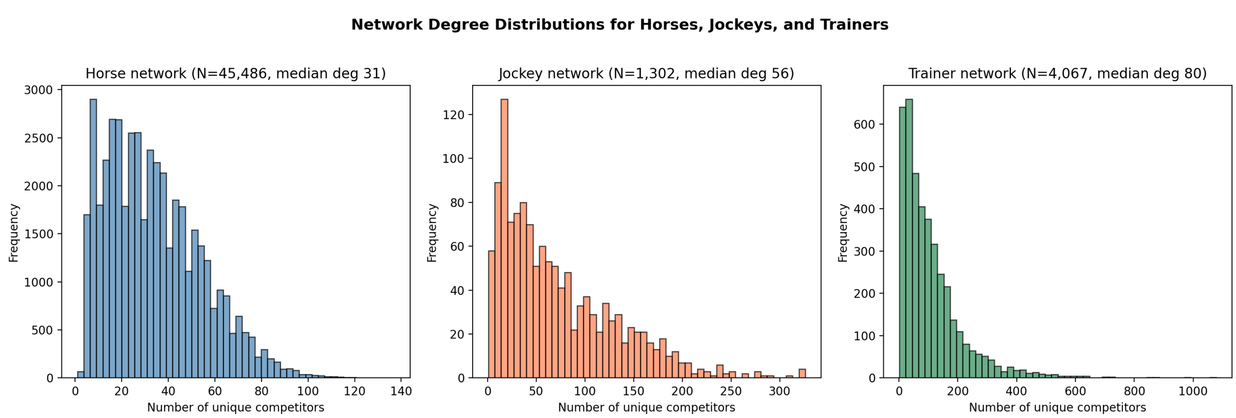


Figure A4: Network Degree Distributions for Horses, Jockeys, and Trainers

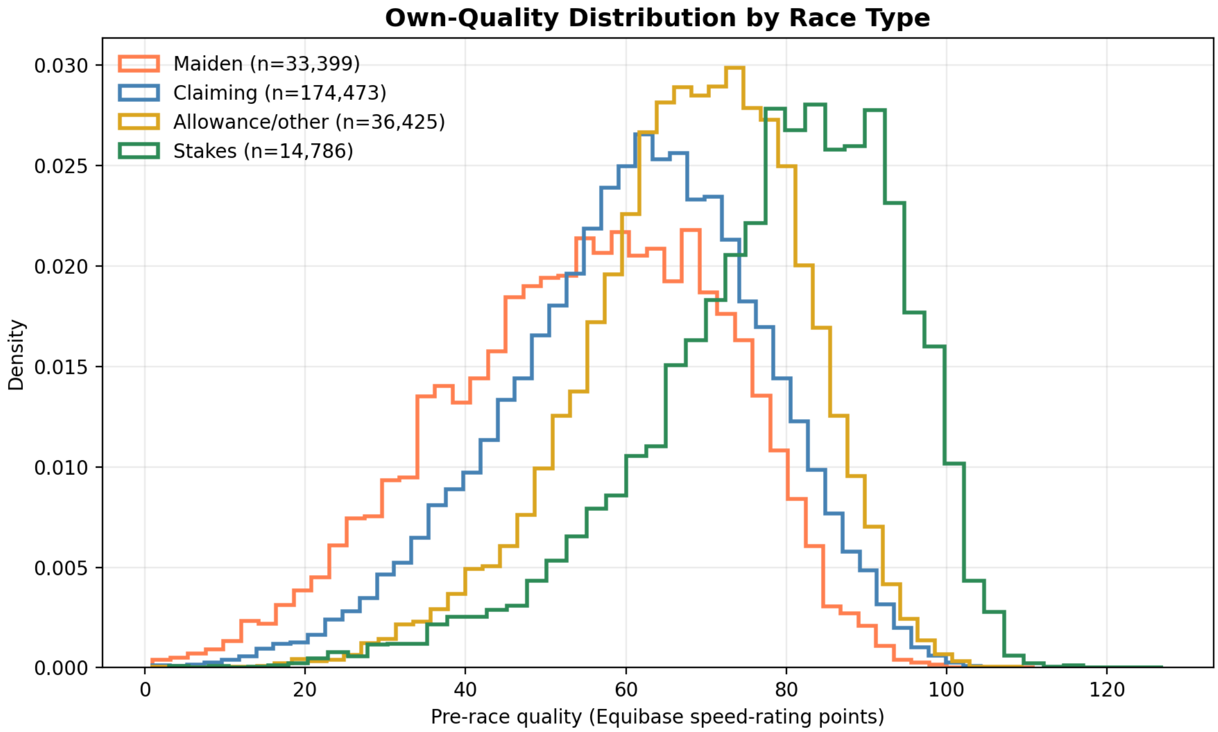


Figure A5: Quality Distributions by Race Type

B. Additional Main Results

Table A2: Familiarity and Peer Effects

	(1) Baseline	(2) Familiarity Levels	(3) Familiarity Interactions
LOO Mean Field Quality	0.226***	0.223***	0.240***
	(0.00668)	(0.00664)	(0.00722)
LOO Mean Jockey Quality	-0.0976***	-0.107***	-0.108***
	(0.0121)	(0.0120)	(0.0120)
LOO Mean Trainer Quality	-0.0677***	-0.0678***	-0.0694***
	(0.00971)	(0.00972)	(0.00972)
Jockey Rival Familiarity (avg prior meetings)		0.0130***	0.0300***
		(0.000935)	(0.00349)
Trainer Rival Familiarity (avg prior meetings)		0.0265***	0.0512***
		(0.00352)	(0.0149)
Field Quality x Jockey Familiarity			-0.000250***
			(0.0000489)
Field Quality x Trainer Familiarity			-0.000329
			(0.000203)
Observations	251408	251408	251408

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) Dirt	(2) Sprint	(3) Claiming	(4) Maiden	(5) Jockey Q	(6) Familiarity	(7) Experience
LOO Mean Field Quality	0.149***	0.196***	0.217***	0.234***	0.178***	0.191***	0.204***
	(0.00940)	(0.00595)	(0.00744)	(0.00658)	(0.00574)	(0.00570)	(0.00564)
Field Quality x Dirt Surface	0.0132						
	(0.00867)						
Field Quality x Sprint (<6f)		-0.00707					
		(0.00613)					
Field Quality x Claiming Race			-0.0328***				
			(0.00626)				
Field Quality x Maiden Race				-0.115***			
				(0.00857)			
Field Quality x Jockey Quality (std)					-0.00432		
					(0.00330)		
Field Quality x Jockey Familiarity (std)						-0.0173***	
						(0.00245)	
Field Quality x Prior Starts (std)							0.0247***
							(0.00507)
Observations	256350	256350	256350	256350	255077	256350	256350

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	0.193***	0.193***	0.201***	0.236***
	(0.00573)	(0.00575)	(0.00611)	(0.00851)
Pre-race Quality (own)	-3.632***	-3.628***	-3.622***	-3.538***
	(0.0336)	(0.0336)	(0.0352)	(0.0476)
Observations	256350	255910	230289	128384

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	0.415***	0.481***	0.346***		0.403***
	(0.00494)	(0.00656)	(0.00804)		(0.00478)
Pre-race Quality (own)	-4.701***	-4.334***	-4.987***	-3.415***	-4.822***
	(0.0356)	(0.0500)	(0.0533)	(0.0420)	(0.0356)
lagged_field_quality				0.0993***	
				(0.00519)	
LOO Mean Field Quality					0.146***
					(0.00557)
Observations	213514	141175	63340	212856	213514

Horse FE in all specs. Clustered by race.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Convexity Within Claiming Subsample (Within-Claim Quality Quartiles)

	(1) Q1 (low)	(2) Q2	(3) Q3	(4) Q4 (high)
field_quality_loo_mean	0.311***	0.339***	0.349***	0.380***
	(0.015)	(0.014)	(0.014)	(0.015)
quality_pre	-3.402***	-3.510***	-3.866***	-3.950***
	(0.072)	(0.099)	(0.108)	(0.115)
Observations	41424	41870	41723	41463

Standard errors in parentheses
Each column: separate horse-FE regression on claiming races within the indicated own-quality quartile.
Quartiles defined on the within-claim distribution of quality_pre (not full-sample quartiles).
Standard errors two-way clustered by race and horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Network-Clustered vs Race-Clustered Standard Errors

	(1) Horse FE (race)	(2) Horse FE (network)	(3) Joint (race)	(4) Joint (network)
LOO Mean Field Quality	0.193***	0.193***	0.226***	0.226***
	(0.00573)	(0.0192)	(0.00668)	(0.0171)
Pre-race Quality (own)	-3.632***	-3.632***	-3.725***	-3.725***
	(0.0336)	(0.0969)	(0.0336)	(0.0918)
LOO Mean Jockey Quality			-0.0976***	-0.0976***
			(0.0121)	(0.0210)
LOO Mean Trainer Quality			-0.0677***	-0.0677***
			(0.00971)	(0.0121)
Observations	256350	256350	251408	251408

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 A8: Network IV Variance Decomposition

Absorbing fixed effect	R^2	Within-share of variance
Track	0.266	0.734
Track × race date	0.276	0.724

The R^2 column reports the share of variance in `indirect_competitor_quality` absorbed by each fixed-effect specification. A large share absorbed by track or track–date FE indicates that the instrument primarily captures between-track or between-meet variation rather than race-specific variation, consistent with the Angrist (2014) concern that network-based instruments can behave as group dummies.

Table A9: Tripartite Anderson-Rubin Pre-Test

	(1) Ranked Peers Only	(2) LOO Means + Ranked Peers
peer_q_rank1	0.0395***	-0.00367
	(0.00837)	(0.00889)
peer_q_rank2	0.0577***	0.00418
	(0.0118)	(0.0124)
peer_q_rank3	0.0990***	-0.0149
	(0.00972)	(0.0118)
peer_jq_rank1	-0.0646***	-0.0634***
	(0.0126)	(0.0131)
peer_jq_rank2	0.00577	0.000764
	(0.0178)	(0.0185)
peer_jq_rank3	-0.0431***	-0.0367**
	(0.0148)	(0.0176)
peer_tq_rank1	-0.0112	-0.00721
	(0.00832)	(0.00885)
peer_tq_rank2	-0.0179	-0.0150
	(0.0118)	(0.0123)
peer_tq_rank3	-0.0289***	-0.0197*
	(0.00995)	(0.0118)
LOO Mean Field Quality		0.242***
		(0.0155)
LOO Mean Jockey Quality		-0.0134
		(0.0223)
LOO Mean Trainer Quality		-0.0315*
		(0.0167)
Observations	250188	250188

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 A10: Entry Model: Scratch Decisions and Field Quality

	(1) Any Scratch	(2) Strategic (A)	(3) Interaction	(4) Claiming	(5) Non-Claiming
main					
Expected LOO Field Quality	0.00532***	0.00787***	0.0413***	-0.00681***	0.0241***
	(0.00170)	(0.00185)	(0.00343)	(0.00256)	(0.00272)
Own Quality (avg pre-race)	-0.00896***	-0.00806***	0.0246***	-0.00265	-0.0125***
	(0.00129)	(0.00149)	(0.00332)	(0.00219)	(0.00204)
Entered Field Size	0.160***	0.175***	0.174***	0.182***	0.166***
	(0.00216)	(0.00264)	(0.00266)	(0.00361)	(0.00396)
purse	-0.000000799***	-0.000000517***	-0.000000227*	-0.00000246***	-0.000000229*
	(0.000000128)	(0.000000132)	(0.000000128)	(0.000000868)	(0.000000137)
distance	0.000665***	0.0000143	0.0000221	0.000762***	-0.000497***
	(0.0000747)	(0.0000886)	(0.0000898)	(0.000149)	(0.000120)
Expected Field Quality x Own Quality			-0.000552***		
			(0.0000493)		
Observations	490432	490432	490432	325385	165047

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

Table A11: Race-Level Scratch Intensity and Field Quality

	(1) All Scratches	(2) A-type	(3) A-type (Claiming)	(4) A-type (Non-Claiming)
Race Mean Quality	0.00269**	-0.00718***	-0.00652***	-0.00793***
	(0.00104)	(0.000692)	(0.000825)	(0.00133)
(first) purse	-0.000000610*	-0.000000464***	-0.000000928	-0.000000593***
	(0.000000366)	(0.000000155)	(0.000000849)	(0.000000175)
(first) field_size	-0.0768***	0.244***	0.212***	0.296***
	(0.00921)	(0.00576)	(0.00674)	(0.0107)
Observations	36495	20763	13853	6910

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

D. Extensions

Table A12: 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	0.193***	0.188***	0.331***	0.0294**	0.226***
	(0.00573)	(0.00571)	(0.0129)	(0.0136)	(0.00668)
Field Quality x Eigenvector Centrality		0.644***			
		(0.113)			
Field Quality x Clustering Coefficient			-0.581***		
			(0.0473)		
Field Quality x PageRank				0.509***	
				(0.0402)	
LOO Mean Jockey Quality					-0.0976***
					(0.0121)
LOO Mean Trainer Quality					-0.0677***
					(0.00971)
Observations	256350	256350	256350	256350	251408

Cols 1-4: Horse FE, clustered by race. Col 5: Horse+Jockey+Trainer FE, two-way clustered. Centrality levels absorbed by agent FE; interactions reported where non-collinear.

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

Table A13: Temporal Dynamics: First vs Second Half of Year

	(1) H1 Horse FE	(2) H2 Horse FE	(3) H1 Tripartite	(4) H2 Tripartite	(5) Pooled Interaction
LOO Mean Field Quality	0.127***	0.129***	0.150***	0.173***	0.230***
	(0.00792)	(0.00788)	(0.00936)	(0.00933)	(0.00653)
Pre-race Quality (own)	-6.068***	-3.865***	-6.131***	-3.884***	-3.673***
	(0.0580)	(0.0509)	(0.0602)	(0.0512)	(0.0337)
LOO Mean Jockey Quality			-0.0595***	-0.128***	
			(0.0159)	(0.0179)	
LOO Mean Trainer Quality			-0.0577***	-0.0415***	
			(0.0122)	(0.0151)	
Field Quality x H2					-0.0497***
					(0.00557)
Second Half of Year					4.623***
					(0.369)
Observations	116159	133209	111712	132438	256350

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 A14: Jockey Switching (v2: Fixed Pre-2023 Baseline)

	(1) Levels	(2) Levels+Fam	(3) FD	(4) FD+Controls	(5) Interaction
Delta Jockey Baseline (v2: fixed pre-2023)	0.0565***	0.0487***	0.0246***	0.0338***	0.0935***
	(0.00454)	(0.00462)	(0.00765)	(0.00774)	(0.0191)
quality_pre	-3.746***	-3.780***			-3.925***
	(0.0564)	(0.0567)			(0.0566)
jockey_rival_familiarity		0.00832***		-0.00749***	
		(0.000917)		(0.000818)	
days_since_last_race				0.00919***	
				(0.00182)	
field_quality_loo_mean					0.189***
					(0.00806)
Delta Jockey Baseline (v2) x Field Quality					-0.000701**
					(0.000298)
Observations	109652	109652	117613	117613	109652

Δ jockey quality v2 = baseline[current jockey] − baseline[prev jockey], where baseline is the jockey's jockey_quality_pre at first 2023 appearance. Structurally zero in non-switch sample (falsification target is satisfied by design rather than by an additional regression; see manuscript Section 7). G1-G2, G6: Horse FE, clustered by horse. G3-G4: First-difference, clustered by horse.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E. Advanced Methods

Table A15: Causal Forest Heterogeneous Treatment Effects (Claiming Subsample)

Panel A: Treatment Effect and Calibration				
	Estimate	SE	p-value	
Average treatment effect (ATE)	0.307***	(0.005)		$N = 165{,}520$
Mean forest prediction	0.991	(0.018)		should ≈ 1
Differential forest prediction	1.418	(0.048)	<0.001	rejects homogeneity
Panel B: Distribution of Conditional Effects (CATE)				
	Mean	SD	P25	P75
CATE	0.311	0.111	0.237	0.383
Panel C: Variable Importance (Top 10)				
Rank	Variable	Importance		
1	Horse clustering coefficient	0.222		
2	Distance	0.178		
3	Field size	0.145		
4	Horse eigenvector centrality	0.105		
5	Pre-race quality (own)	0.074		
6	Field quality LOO max	0.055		
7	Pre-race starts	0.041		
8	Age	0.031		
9	Jockey–rival familiarity	0.028		
10	Trainer–rival familiarity	0.022		

Note: Estimated using the `grf` package on the claiming subsample. Treatment is leave-out mean field quality; horse fixed effects absorbed via residualization. The differential forest prediction rejects treatment-effect homogeneity.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Convexity and Supermodularity Tests

Panel A: Polynomial Specification				
Term	Estimate	SE	p-value	Interpretation
Field quality (linear)	0.209***	(0.005)		
Field quality (squared)	0.003***	(0.000)	<0.001	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	19.2		<0.001	Significant nonlinear interaction
Panel C: Decile-Specific Peer Effects				
Decile	Estimate	SE	Mean Quality	
D1	0.125***	(0.016)	30.3	
D2	0.159***	(0.015)	44.2	
D3	0.185***	(0.015)	51.4	
D4	0.211***	(0.014)	56.8	
D5	0.246***	(0.015)	61.2	
D6	0.258***	(0.014)	65.3	
D7	0.263***	(0.015)	69.5	
D8	0.284***	(0.014)	73.9	
D9	0.322***	(0.015)	79.3	
D10	0.317***	(0.016)	88.8	

Note: All specifications include horse fixed effects and race-type controls (claiming/maiden/stakes). N = 259,048. Decile 1 is the lowest own-quality decile; columns report the within-decile horse-FE peer effect.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Partial Network Bias Correction

Panel A: Analytical Correction ($\hat{\beta}/\rho$)

ρ	Corrected Estimate	
0.1	1.928	
0.3	0.643	
0.5	0.386	
0.7	0.275	
1.0	0.193	(observed)

Panel B: Simulation-Based Correction (50 Simulations)

ρ	Corrected Estimate	Attenuation	
0.1	9.919	1.9%	
0.3	2.738	7.0%	
0.5	1.259	15.3%	
0.7	0.622	31.0%	
1.0	0.193	—	(no correction)

Note: ρ denotes the fraction of the full competition network observed. Our one-year sample corresponds to roughly $\rho = 0.3$. The corrected columns are *illustrative*: they indicate that the observed estimate is a lower bound on the full-network effect, but the simulation-based extrapolation (Panel B) is unstable at small ρ—the $\rho = 0.1$ row in particular should not be read as a literal magnitude. The robust takeaway is qualitative: partial observation of the network attenuates the estimate toward zero.

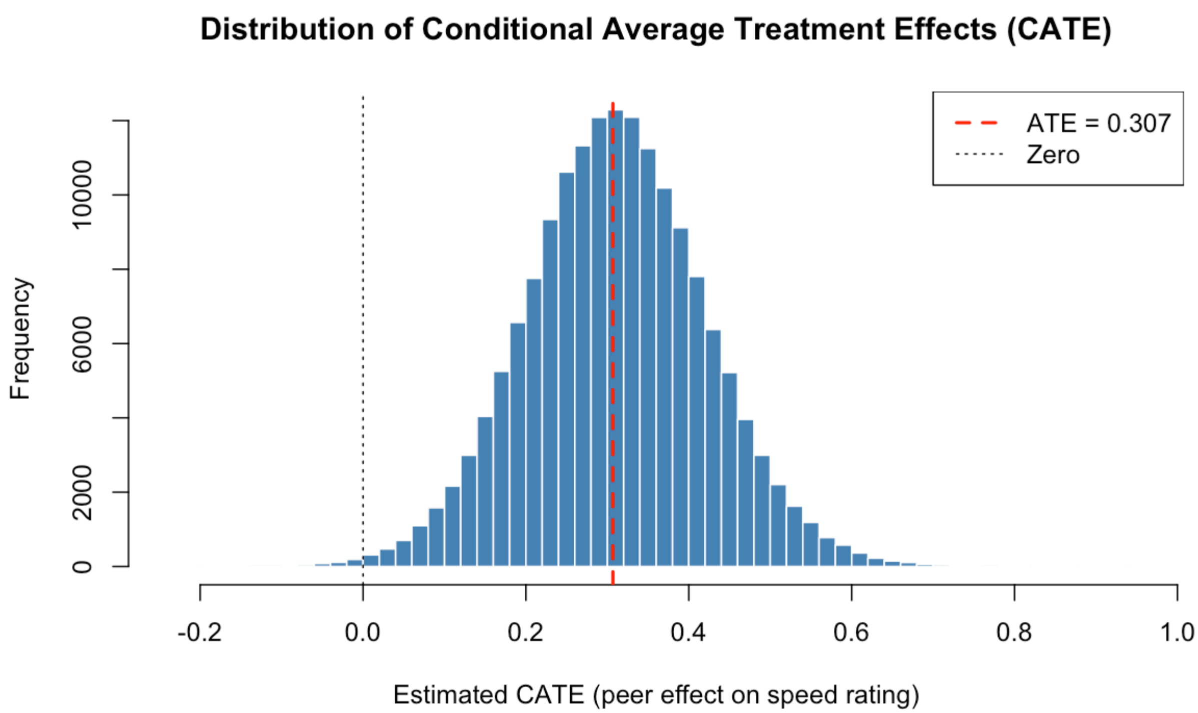


Figure A6: Distribution of Conditional Average Treatment Effects from Causal Forest

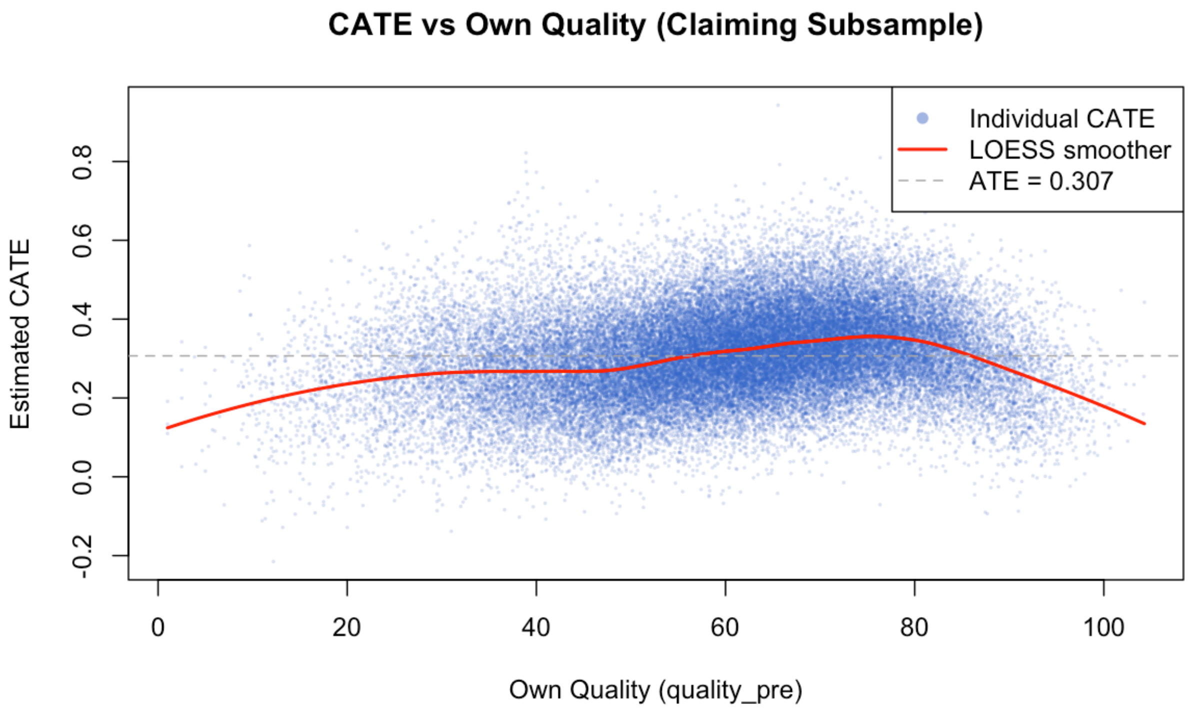


Figure A7: Causal Forest Conditional Average Treatment Effect (CATE) Estimates by Horse Quality Decile

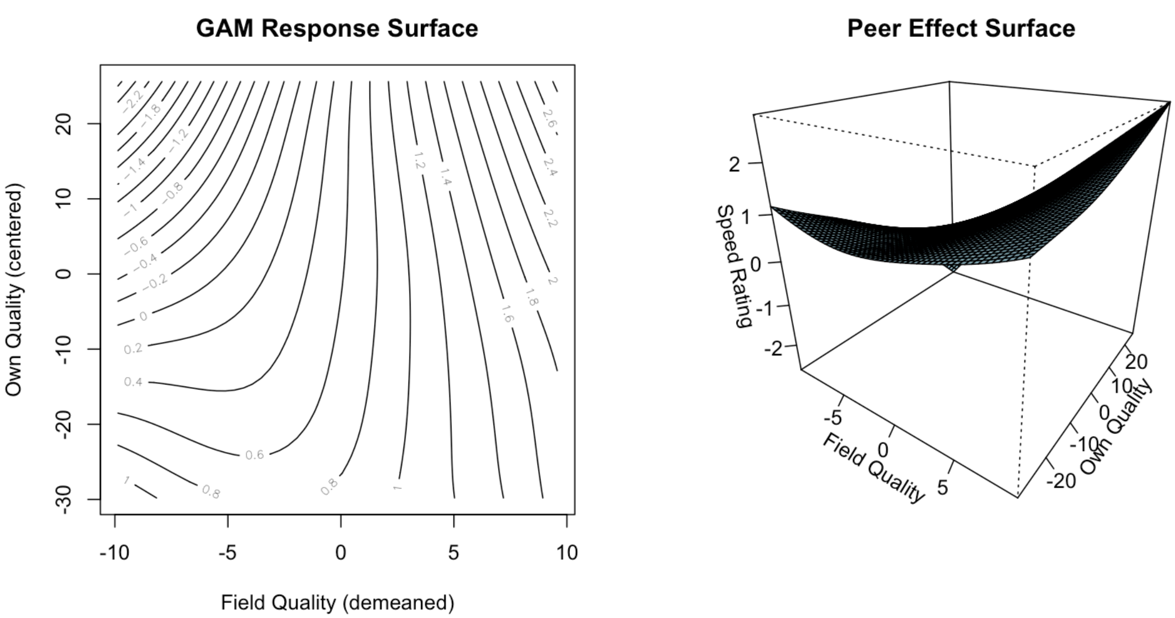


Figure A8: Convexity of Peer Effects: Response Surface over Own and Field Quality

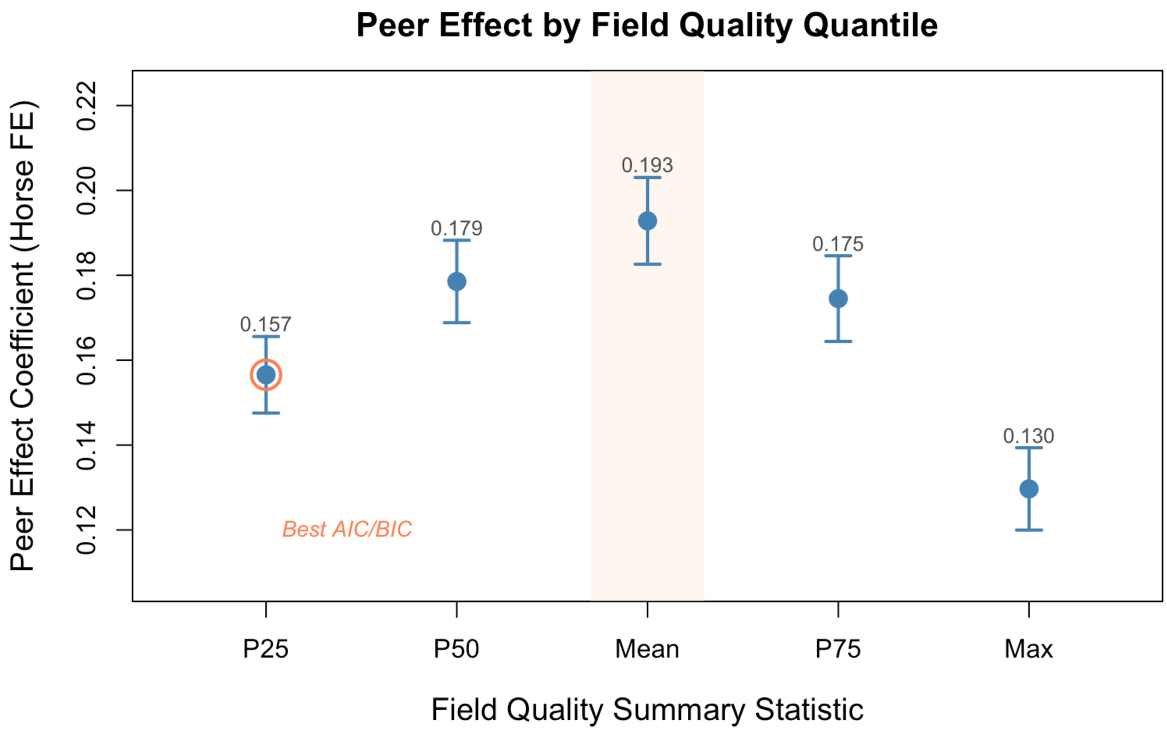


Figure A9: Peer Effect Coefficients across the Conditional Speed Rating Distribution

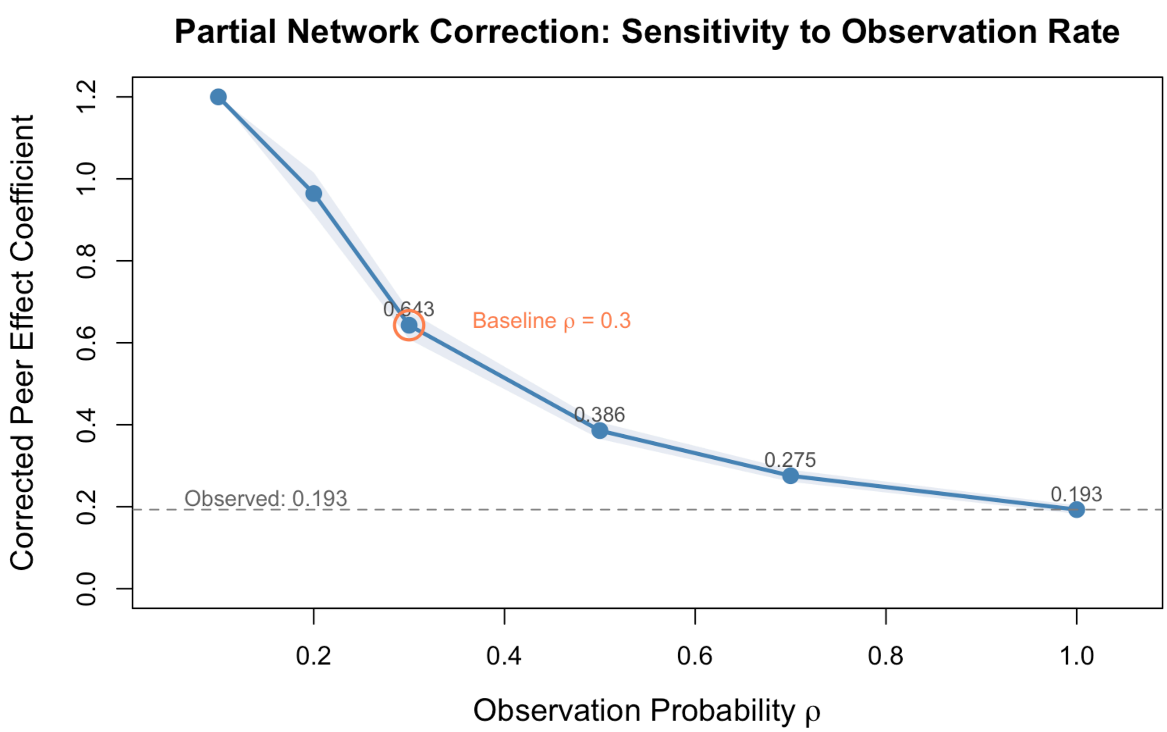


Figure A10: Sensitivity of Peer Effect Estimates to Partial Network Observation

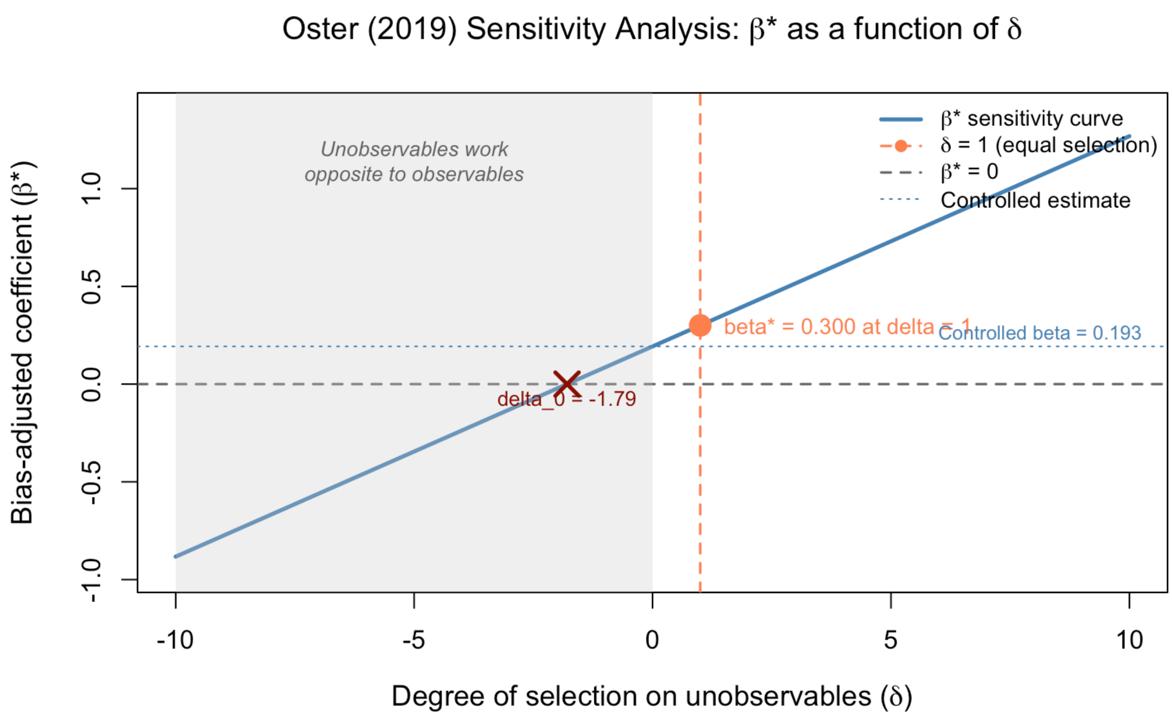


Figure A11: Oster (2019) Bounds on Peer Effect Estimates under Proportional Selection

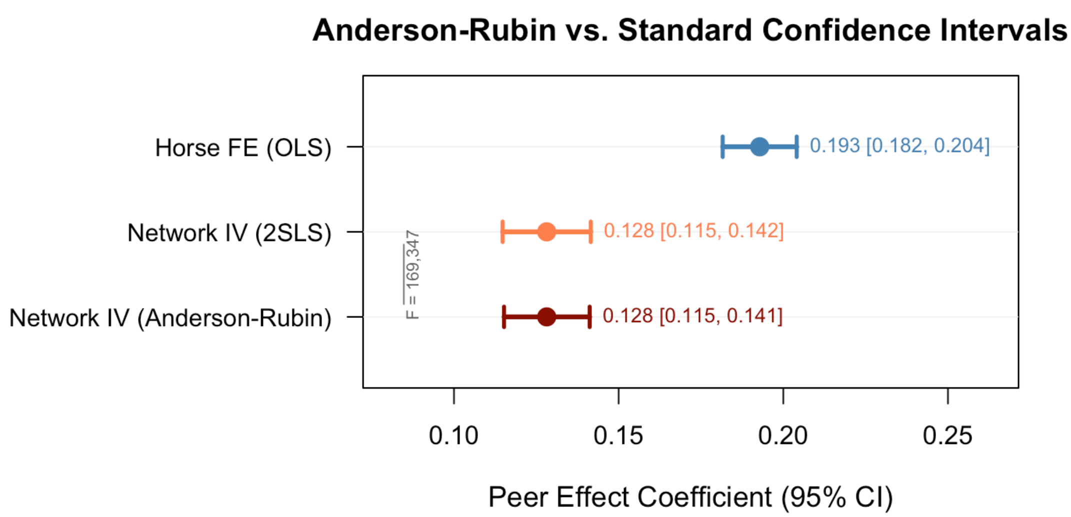


Figure A12: Anderson–Rubin Confidence Intervals vs. Conventional IV Inference

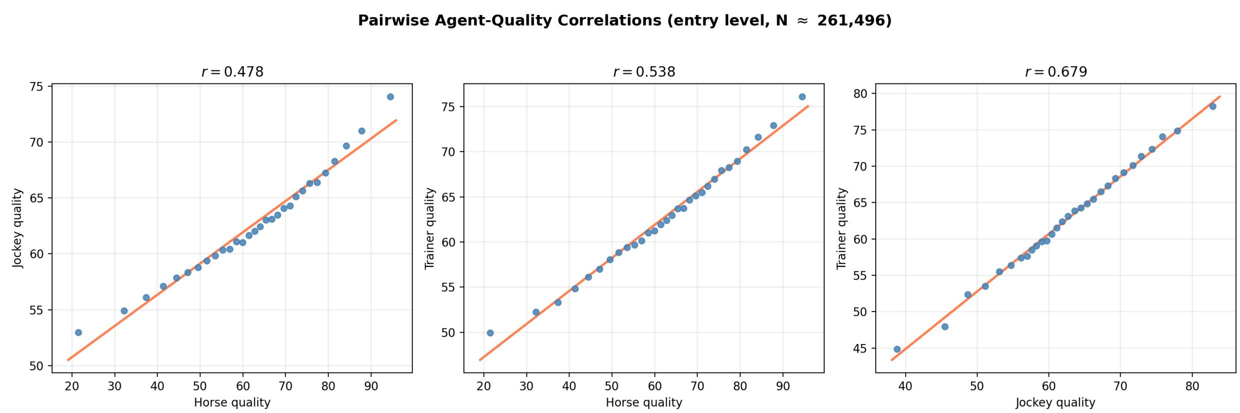


Figure A13: Pairwise correlations among horse, jockey, and trainer quality measures (entry-level scatter, $N \approx 260{,}000$). Reported correlations: $r = 0.478$ (horse–jockey), $r = 0.538$ (horse–trainer), $r = 0.679$ (jockey–trainer). The strong assortative-matching pattern motivates the orthogonalization robustness check in Section 7.3.

Figure A14: Within-horse demeaned binned scatter of speed rating against field quality, separately within each own-quality quartile. Slopes rise monotonically: 0.229 (Q1, lowest), 0.291 (Q2), 0.310 (Q3), 0.328 (Q4, highest), confirming the convex pattern reported in Section 7.4 as a scatter companion to the linear estimates.

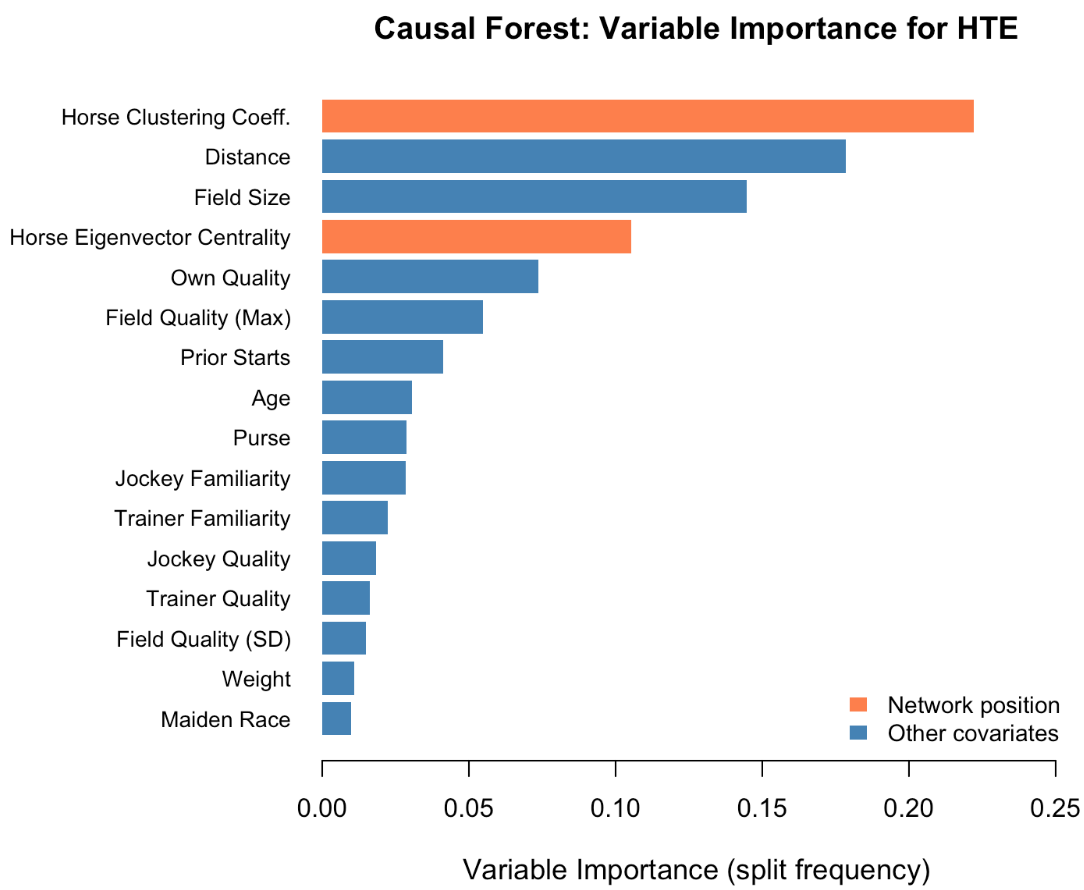


Figure A15: Causal forest variable importance for the heterogeneous treatment effect of field quality. The horse's network clustering coefficient is the single highest-ranked feature, and eigenvector centrality also ranks among the top covariates, indicating that competitive-network structure shapes which horses respond more strongly to peer quality.

F. Data Attribution and Disclaimer

The analyses in this paper use data provided by Equibase Company LLC under the Free Dataset program. The following attribution and disclaimer are reproduced at Equibase's request.

© 2026 Equibase Company LLC, all rights reserved. The data used to create the report are proprietary to and are copyrighted by Equibase Company LLC. Unauthorized copying of such data, including modifications thereof, or the creation of programs or products in which such data have been merged or included with other data programs or products, for distribution to third parties, whether gratuitously or for sale, is expressly forbidden. You may be held legally responsible for any proprietary or copyright infringement that is caused or encouraged by your failure to comply with these Terms and Conditions of Use.

Equibase shall not have any responsibility or liability arising from the use of the information, products, programs or services furnished hereunder.

Data provided or compiled by Equibase Company LLC generally are accurate, but occasionally errors and omissions occur as result of incorrect data received by others, mistakes in processing and other causes. Therefore, Equibase Company LLC makes no representations or warranties, expressed or implied, including those of merchantability, accuracy, timeliness or fitness for a particular purpose, as to any matter whatsoever, including, but not limited to, the accuracy of any information, product, program or service furnished hereunder and you the buyer accept any and all such information, products, programs, and/or services on an "as is" basis.