



UTMD Working Paper

The University of Tokyo
Market Design Center

UTMD-106

Team for Speed:

Nonparametric Evidence on Heterogeneous Skill-Specific Affinity in Team Production

Masaya Nishihata

Keio University

Mitsubishi UFJ Research and Consulting Co., Ltd.

Suguru Otani

University of Tokyo

December 24, 2025

Team for Speed: Nonparametric Evidence on Heterogeneous Skill-Specific Affinity in Team Production

Masaya Nishihata*, Suguru Otani†

December 24, 2025

Abstract

We examine whether team affinity differs across skill dimensions in team production. Using a novel nonparametric framework that accommodates task-level structure, role asymmetry, and latent affinity, we decompose team performance into skill-specific productivity and unobserved match affinity. As an illustrative application, we analyze elite women’s bobsleigh data, where performance can be separated into start and riding phases with distinct individual skill inputs. The estimates reveal heterogeneous, task-specific affinities: coordination and complementarity are stronger in the start phase but weaker and more dispersed during riding, underscoring skill-specific heterogeneity in unobserved team affinity.

Keywords: team production, production function, multidimensional skills, nonparametric estimation

JEL code: J24, D24, C14

1 Introduction

How teams transform individual abilities into collective outcomes is a central question in the economics of team production. Classical models of team production functions often assume a representative agent, symmetric roles, or fixed task assignments, abstracting from the complexity introduced by task-level specialization, heterogeneous skills, and role asymmetry. In reality, however, most productive activity involves teams composed of individuals with distinct abilities and

*nishihata-masaya@keio.jp, Graduate School of Economics, Keio University and Mitsubishi UFJ Research and Consulting Co., Ltd.

†suguru.otani@e.u-tokyo.ac.jp, Department of Economics, The University of Tokyo

‡This work was supported by JST ERATO JPMJER2301, 25K16620, and JST SPRING JPMJSP2123.

responsibilities, and how those abilities interact—both technically and behaviorally—can significantly shape overall performance (Deming 2023).

Despite these complexities, most theoretical and empirical studies continue to rely on unidimensional proxies for ability. This limits our ability to evaluate how human capital should be optimally assigned or invested in real-world settings where tasks are separable, roles are asymmetric, and individuals bring multidimensional skill sets to the table. In particular, little is known about whether the degree of match efficiency or affinity between teammates differs systematically across skill dimensions or whether complementarities are homogeneous across tasks. Understanding the heterogeneity of such affinities is crucial for assessing how skill composition and coordination jointly determine team output.

We provide new evidence on heterogeneity in team affinity across skill dimensions using a nonparametric framework for multidimensional skills and asymmetric roles, following Matzkin (2003) and Lange and Papageorgiou (2020).¹ The framework accommodates task-level structure, role asymmetry, and latent affinity. It estimates task-specific production functions and latent match affinities without functional-form restrictions, exploiting variation in teammate assignments to decompose performance into skill-specific productivity and unobserved affinities capturing within-task coordination.

Our application exploits two-woman bobsleigh data with a clean two-task structure (start and riding), precise performance measures, and strict driver–brakeman role separation. These features let us isolate skill-specific affinity rather than conflate it with role sorting or aggregate ability. We find substantial heterogeneity in team affinity across dimensions: start-phase affinity shows structured high–high clusters with additional efficient off-diagonal pairs, while riding affinity is weaker, more diffuse, and often slow even among high-skill athletes, with little diagonal gradient or stability across attempts. The contrast shows affinity is task-specific and that coordination beyond measurable skill meaningfully shapes performance.

We relate to two strands. First, empirical work on team production studies learning spillovers, complementarities, and peer effects, but typically treats affinity as a single latent factor rather than allowing variation across skill dimensions.² Recent structural work separates individual contributions (Bonhomme 2021, Xu *et al.* 2024) yet often abstracts from explicit tasks or separable roles. Our framework nonparametrically decomposes performance into skill-specific productivity and skill-specific affinities. Second, literature on multidimensional skills and task-based production (Lindenlaub 2017, Lise and Postel-Vinay 2020, Woessmann 2024, Herkenhoff *et al.* 2024, Bartel

¹Applications include labor markets (Otani 2024, 2025b, Kanayama and Otani 2024), transportation (Brancaccio *et al.* 2020), and marriage markets (Otani 2025a).

²Examples include Jarosch *et al.* (2021), Herkenhoff *et al.* (2024), Jäger and Heining (2022), and peer effects studies (Falk and Ichino 2006), Mas and Moretti (2009), Bandiera *et al.* (2010). For sports, see Arcidiacono *et al.* (2017) and Cohen-Zada *et al.* (2024). See also De Loecker and Syverson (2021) on production function estimation.

et al. 2014) usually relies on coarse proxies; our setting provides direct running and riding measures, enabling estimation of task-specific affinity conditional on partner skill and revealing heterogeneity in teamwork ability (Deming 2017, Weidmann and Deming 2021).

2 Model

We develop a framework for team production with multiple tasks and asymmetric roles. Teams combine (i) additively separable task-specific skills and (ii) unobserved within-task affinity converting skills nonadditively into performance. We do not impose optimizing behavior because formation reflects preferences, managerial discretion, and other frictions. The formulation covers two-person teams splitting responsibilities (analytical vs. execution, technical vs. managerial). Our bobsleigh application exploits clean task separation and precise measurement.

For a two-person team ℓ at time t performing K tasks ($k = 1, \dots, K$), let $X_{\ell kt}$ and $Y_{\ell kt}$ denote the leader’s (Player 1) and assistant’s (Player 2) skill inputs. Output on task k is

$$H_{\ell kt} = m_k(A_{\ell kt}X_{\ell kt}, Y_{\ell kt}),$$

where $A_{\ell kt}$ denotes unobserved team–task-specific efficiency—hereafter used synonymously with affinity—entering the production function nonadditively, and $m_k(\cdot, \cdot)$ is the task-specific production function.

Assume (i) constant returns to scale (CRS), yielding $H_{\ell kt}/X_{\ell kt} = m_k(A_{\ell kt}, Y_{\ell kt}/X_{\ell kt})$, and (ii) conditional independence between $A_{\ell kt}$ and $X_{\ell kt}$ given $Y_{\ell kt}$. CRS permits arbitrary forms beyond Cobb–Douglas. Conditional independence posits leader skill is uncorrelated with latent affinity conditional on assistant skill—plausible under institutional assignment, frequent partner switching, or managerial discretion rather than chemistry-based sorting.³

Nonparametric identification and estimation of $m_k(\cdot, \cdot)$ and latent efficiency $A_{\ell kt}$ follow Matzkin (2003) and Lange and Papageorgiou (2020). Given sufficient variation in teammate assignments and observed task performance, and under CRS and conditional independence, m_k is identified up to scale, together with the distribution of $A_{\ell kt}$. Observing $(X_{\ell kt}, Y_{\ell kt}, H_{\ell kt})$ across teams and periods allows nonparametric recovery of $A_{\ell kt}$ from variation in $H_{\ell kt}$ conditional on $Y_{\ell kt}$. See Appendix A for estimation details.

³Identification requires conditional independence; pervasive violations would undermine identification. Nationality rules and an institutionally restricted choice set limit feasible matches, so within-nationality sorting on latent affinity is likely local rather than systematic in the overall pool. Appendix A.1.4 documents dispersed pairings consistent with limited sorting. If the assumption fails, the estimates will be biased. An instrumental-variables approach could help (Imbens and Newey 2009), although suitable instruments are difficult to find.

3 Data

We use data from two-woman bobsleigh competitions. The setting is ideal because performance splits cleanly into start and riding tasks, driver–brakeman roles are sharply asymmetric, and standardized international records generate rich formation variation. The sport provides a tractable, data-rich case with transparent links between inputs (running and riding skills) and outputs (task times).

We compile official IBSF race results from 2015 to 2025 (accessed May 5, 2025), covering digitized two-attempt two-woman and monobob events including World Cups and Olympics.⁴ Some competitions have more than two attempts; we retain the first two for comparability (about 11% exceed two), and excluding them leaves patterns unchanged. Rankings depend on cumulative times, so strategic considerations related to run counts are limited. IBSF standardization of timing, weight, and equipment minimizes non-skill variation and ensures comparability across tracks. For each run we collect start, finish, and split times plus athlete identities, nationalities, and starting orders. Race-level regressions with race and player fixed effects reconstruct solo performance and team performance, absorbing track and weather effects. Individual skill comes from monobob fixed effects, team performance from residualized team race times; both are shifted to positive values for the production domain. The sample includes 45 athletes, 47 unique matches, and 160 team runs, and requires recorded solo monobob performances for skill identification; partner rotation produces dispersed pairings used in estimation.

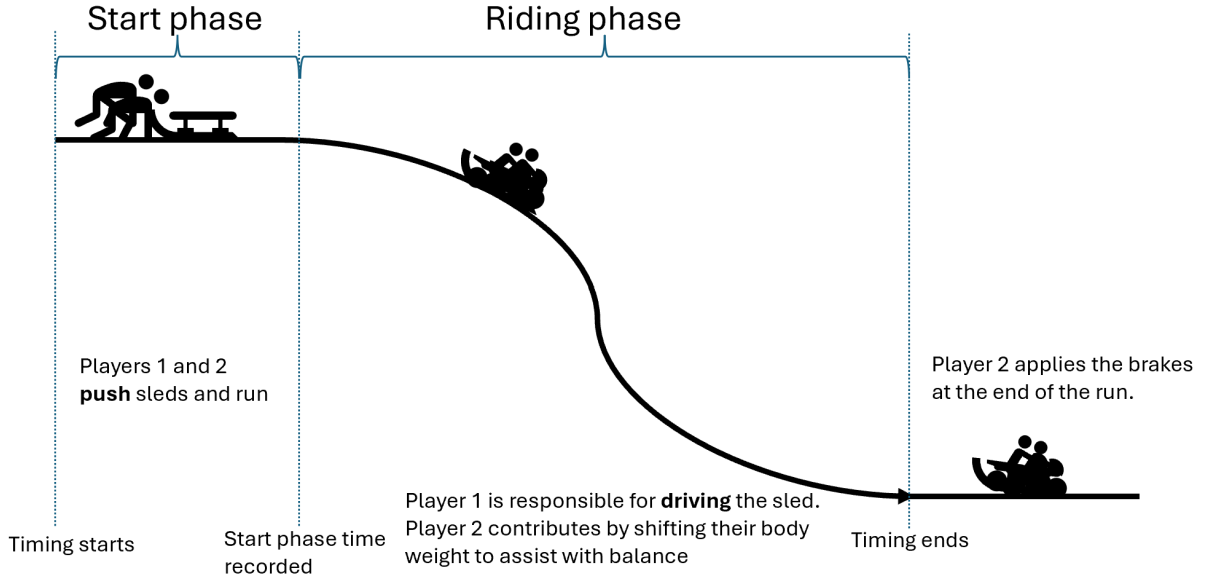


Figure 1: Bobsleigh Competition Design

Bobsleigh races have two phases—start and riding—emphasizing distinct skills (Figure 1). Start

⁴<https://www.ibsf.org/en/races-and-results>

performance reflects explosive power and synchronization, and team affinity affects how smoothly athletes load into the sled; riding depends on steering and balance along the track. In monobob, one athlete performs all roles; in two-woman events, the driver steers while the brakeman balances and brakes. This division isolates physical strength from technical control and mirrors team settings with clearly delineated tasks. Rankings use the sum of two run times, and strict weight and equipment rules limit non-skill variation, so performance mainly reflects ability and coordination, enabling clean identification of task-specific skill and affinity within a unified measurement frame.

Figure 2 presents model-free residualized team performance across terciles of solo performance for each role (1 = lowest, 3 = highest).⁵ Cells report average residualized performance (bright = faster-than-expected, dark = slower). Higher-skill teams generally perform better, but the pattern is neither monotonic nor symmetric: several mixed-skill pairs match or outperform high-high pairs, and some high-high pairs underperform. Structure is broadly stable across attempts yet differs sharply between start and riding phases, indicating that observed skill sorting alone cannot explain outcomes and highlighting unobserved, task-specific affinity beyond additive skill effects.

4 Results

Figure 3 displays implied efficiency between drivers (Player 1) and brakemen (Player 2), with axes ordered by solo-skill fixed effects (Section 3).⁶ Colors show efficiency relative to a baseline (white = 100; red = faster-than-expected; blue = slower). In the start phase, efficient performance clusters among high-high pairs with extra red pockets off the diagonal, while blue regions lie mainly in low-skill cells. Patterns shift across attempts—some pairs remain efficient, others flip to blue—suggesting stochastic elements in push coordination. The wide red-blue dispersion conditional on skill indicates that latent affinity strongly shapes start performance and that high observable skill alone does not guarantee efficient execution. Table 1 reports summary statistics.

Table 1: Summary Statistics of Mean Implied Efficiency

	N	Mean	SD	Min	Max
Start-phase (1st attempt)	47	371.91	141.56	1.00	747.73
Start-phase (2nd attempt)	47	125.31	44.49	1.00	243.18
Riding-phase (1st attempt)	47	147.70	120.85	1.00	646.82
Riding-phase (2nd attempt)	47	158.56	197.43	1.00	979.82

⁵See Appendix A.1.1 for estimation of individual performance via fixed effects.

⁶Lower times imply better performance, so raw times are negated and shifted so higher values mean faster performance.

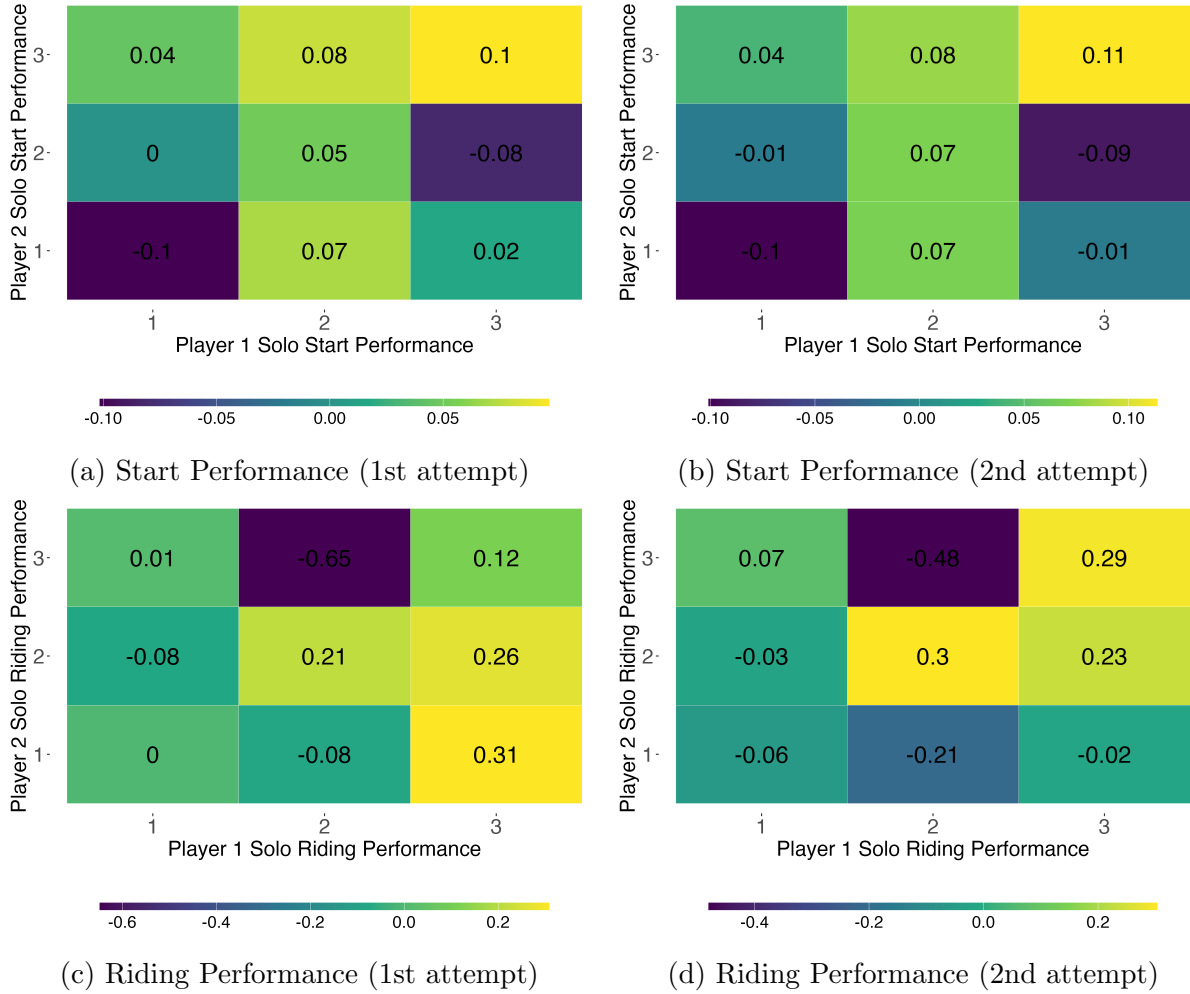


Figure 2: Average Residualized Team Start and Riding Performance Conditional on Individual Performance-Bin Pairs

Note: 1 = lowest performance, 3 = highest performance

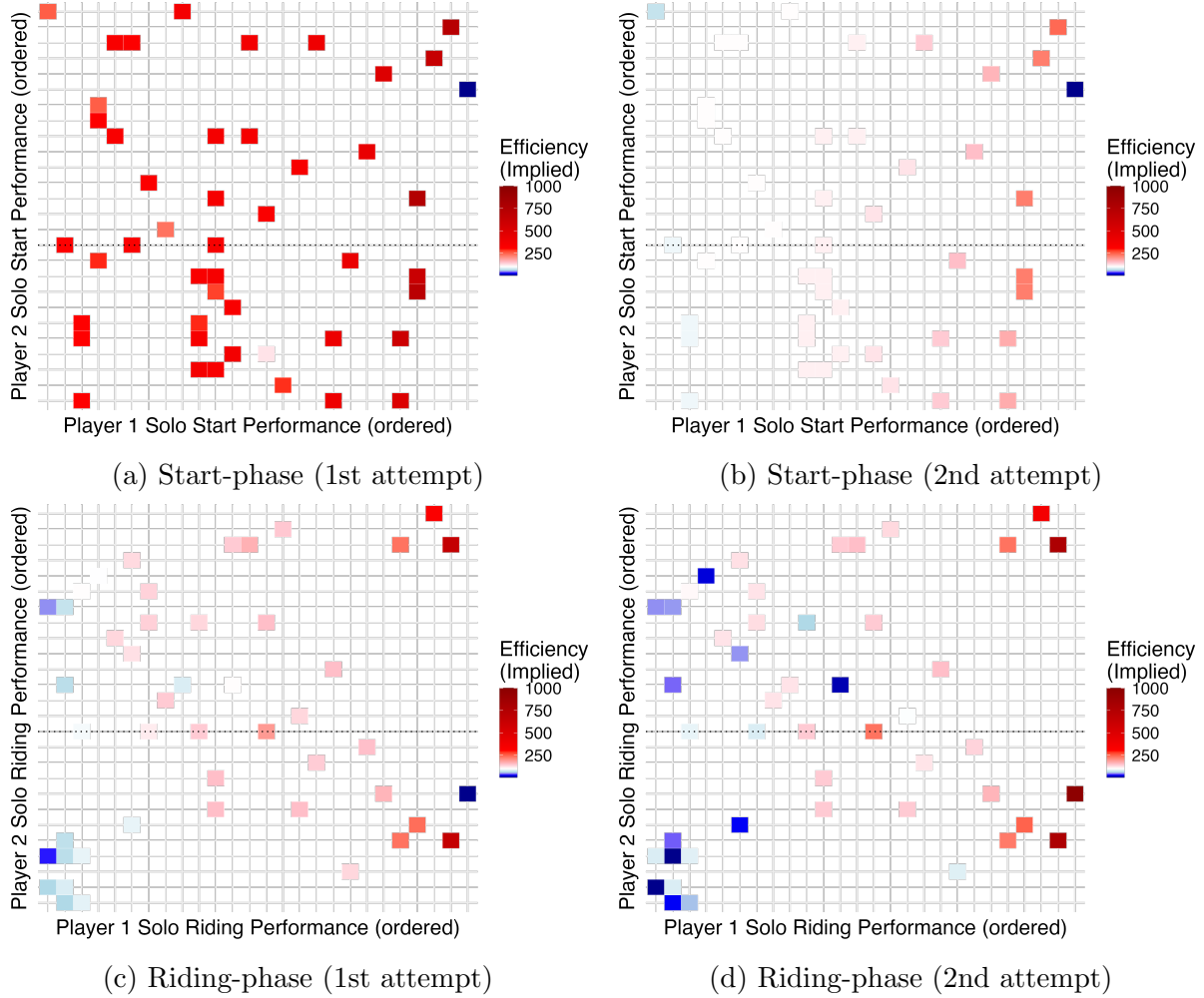


Figure 3: Implied efficiency between Player 1 (driver) and Player 2 (brakeman)

Note: Efficiency is normalized to 100 (white) at the median skill of Player 2 (dotted line). Each axis is ordered by fixed effects estimated from residualized solo start time performance, and the color intensity reflects the extent of the mean unobserved efficiency of the pair, with red indicating greater efficiency.

In the riding phase, efficiency is weaker and diffuse: red cells are sporadic, many combinations—especially high-driver pairings in the second attempt—yield blue (slower) cells, and there is no diagonal gradient. Riding therefore depends on interaction-intensive coordination (steering, balance, line choice) that solo skill does not capture. The broad dispersion, including many slow performances among skilled individuals, underscores pronounced task-specific heterogeneity in affinity.⁷

These patterns align with asymmetric roles. During the start, both athletes push, so affinity reflects coordination in synchronized effort. During riding, the driver steers while the brakeman is largely passive, so affinity captures residual interactions rather than symmetric teamwork. Efficiency gains attributable to Player 2 are modest and weakly aligned across phases, consistent with the dispersed riding-phase affinity.

5 Conclusion

Our analysis shows that team affinity varies across skill dimensions. Using a nonparametric framework that separates task-specific skills from latent efficiency, we find that coordination and complementarities differ across tasks, indicating that observed skill assortativeness alone cannot explain team outcomes. Task-specific affinity is thus a key determinant of collective performance in multidimensional team production.

⁷Each phase’s solo-skill fixed effect decomposes into phase-specific skill and a non-specific component (e.g., general athletic ability). Correlated non-specific skill across phases can generate correlated individual fixed effects (Appendix A.1.3). Thus, high phase affinity may reflect general non-specific affinity, phase-specific matching, or both.

References

- ARCIDIACONO, P., KINSLER, J. and PRICE, J. (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics*, **35** (1), 191–225.
- BANDIERA, O., BARANKAY, I. and RASUL, I. (2010). Social incentives in the workplace. *Review of Economic Studies*, **77** (2), 417–458.
- BARTEL, A. P., BEAULIEU, N. D., PHIBBS, C. S. and STONE, P. W. (2014). Human capital and productivity in a team environment: evidence from the healthcare sector. *American Economic Journal: Applied Economics*, **6** (2), 231–259.
- BONHOMME, S. (2021). Teams: Heterogeneity, sorting, and complementarity. *arXiv preprint arXiv:2102.01802*.
- BRANCACCIO, G., KALOUPTSIDI, M. and PAPAGEORGIOU, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, **88** (2), 657–691.
- COHEN-ZADA, D., DAYAG, I. and GERSHONI, N. (2024). Effort peer effects in team production: Evidence from professional football. *Management Science*, **70** (4), 2355–2381.
- DE LOECKER, J. and SYVERSON, C. (2021). An industrial organization perspective on productivity. In *Handbook of industrial organization*, vol. 4, Elsevier, pp. 141–223.
- DEMING, D. J. (2017). The growing importance of social skills in the labor market. *The quarterly journal of economics*, **132** (4), 1593–1640.
- (2023). Multidimensional human capital and the wage structure. *Handbook of the Economics of Education*, **7**, 469–504.
- FALK, A. and ICHINO, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, **24** (1), 39–57.
- HERKENHOFF, K., LISE, J., MENZIO, G. and PHILLIPS, G. M. (2024). Production and learning in teams. *Econometrica*, **92** (2), 467–504.
- IMBENS, G. W. and NEWEY, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77** (5), 1481–1512.
- JÄGER, S. and HEINING, J. (2022). *How substitutable are workers? evidence from worker deaths*. Tech. rep., National Bureau of Economic Research.

- JAROSCH, G., OBERFIELD, E. and ROSSI-HANSBERG, E. (2021). Learning from coworkers. *Econometrica*, **89** (2), 647–676.
- KANAYAMA, H. and OTANI, S. (2024). Nonparametric estimation of matching efficiency and elasticity in a spot gig work platform: 2019-2023. *arXiv preprint arXiv:2412.19024*.
- LANGE, F. and PAPAGEORGIOU, T. (2020). *Beyond Cobb-Douglas: flexibly estimating matching functions with unobserved matching efficiency*. Tech. rep., National Bureau of Economic Research.
- LINDENLAUB, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, **84** (2), 718–789.
- LISE, J. and POSTEL-VINAY, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, **110** (8), 2328–2376.
- MAS, A. and MORETTI, E. (2009). Peers at work. *American Economic Review*, **99** (1), 112–145.
- MATZKIN, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, **71** (5), 1339–1375.
- OTANI, S. (2024). Nonparametric estimation of matching efficiency and mismatch in labor markets via public employment security offices in japan, 1972-2024. *arXiv preprint arXiv:2407.20931*.
- (2025a). Nonparametric estimation of matching efficiency and elasticity in a marriage agency platform: 2014–2025. *Economics Letters*, **256**, 112617.
- (2025b). Nonparametric estimation of matching efficiency and elasticity on a private on-the-job search platform: Evidence from japan, 2014-2024. *Journal of the Japanese and International Economies*, p. 101394.
- WEIDMANN, B. and DEMING, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, **89** (6), 2637–2657.
- WOESSMANN, L. (2024). Skills and earnings: A multidimensional perspective on human capital. *Annual Review of Economics*, **17**.
- XU, Y., HU, M., CHU, J. and CHING, A. T. (2024). Heterogeneous complementarity and team design: The case of real estate agents. *Marketing Science*.

A Appendix (Not for publication)

A.1 Data Appendix

A.1.1 Construction of Individual Fixed Effects from Monobob Events

A key input to our analysis is the vector of individual skill fixed effects, which we estimate from solo monobob competition data. This subsection clarifies the estimation procedure and the amount of data used.

Estimation Procedure. Individual skill fixed effects are estimated separately for each performance dimension (start time, finish time, and riding time) using the following regression:

$$Y_{iest} = \alpha_i + \gamma_e + \delta_s + \varepsilon_{it},$$

where Y_{iest} is the performance (start time, finish time, or riding time) for athlete i in event e , race order s , and run t , α_i is the athlete fixed effect capturing individual skill, γ_e is the event (competition) fixed effect absorbing track-specific and weather conditions, and δ_s is the starting order fixed effect controlling for sequencing effects. The athlete fixed effect $\hat{\alpha}_i$ from this regression serves as our measure of individual skill for each dimension. By controlling for event and starting order, we isolate the persistent component of individual performance that reflects inherent ability rather than race-specific conditions.

Our sample for fixed effect estimation covers multiple international monobob events from the IBSF circuit between 2021 and 2025. The dataset contains 729 individual monobob runs across all 135 monobob women athletes. Fixed effects are estimated for all athletes who have participated in at least two monobob competitions.

For the 135 athletes with both monobob records, Figure 4 summarizes the distribution of monobob runs per athlete who has at least two records. The substantial number of runs per athlete provides variation needed to identify individual fixed effects with reasonable precision. Athletes with more runs contribute more precisely estimated fixed effects, while those with fewer runs have larger standard errors. However, even athletes with relatively few monobob runs (e.g., 5–10 runs) provide enough variation to estimate meaningful skill parameters given the inclusion of event fixed effects that absorb much of the run-to-run variance.

The reliability of these fixed effects is supported by multiple observations per athlete, event fixed effects that absorb track and weather variation, and starting-order controls that address sequencing effects. Athletes with fewer monobob runs yield noisier estimates, but event fixed effects reduce residual variance sufficiently for meaningful skill identification.

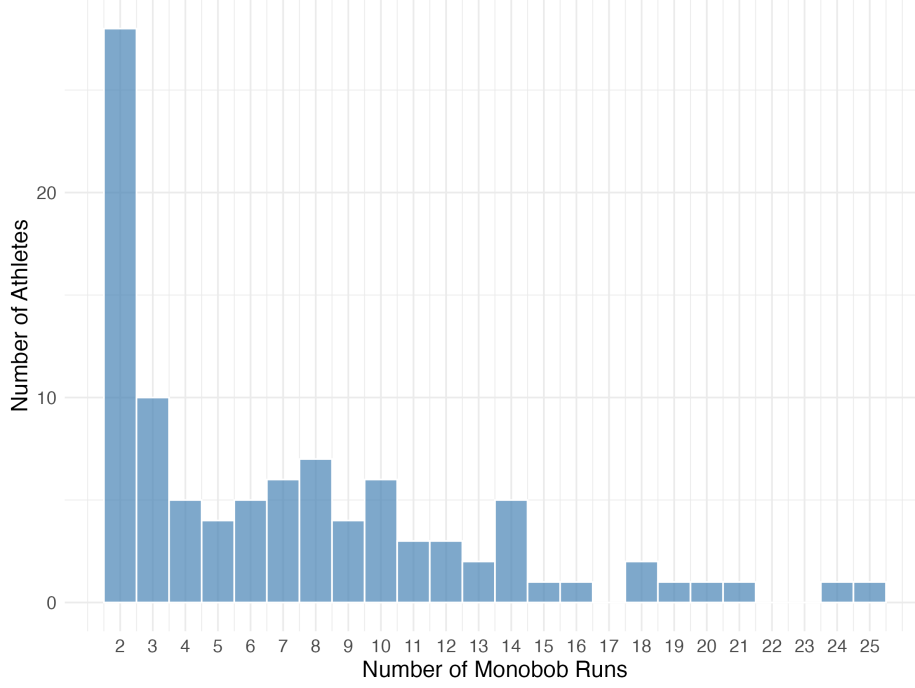


Figure 4: Distribution of Monobob Runs per Athlete

A.1.2 Summary Statistics

Our dataset combines individual- and team-level performance metrics from official records of international two-woman bobsleigh competitions. Race performance is recorded in standardized time units across events and is further disaggregated into the start (running) and riding phases.

Player-level. In Panel (a) of Table 2, we report residualized solo performance for 45 athletes, who are properly merged, along three dimensions: *Residualized Solo Finish Performance*, *Residualized Solo Start Performance*, and *Residualized Solo Riding Performance*. Residuals are obtained by controlling for track-by-event fixed effects and race orders to isolate individual skill components. We then apply a positive-shift transformation so that larger values denote better performance and the support is strictly positive. For any residualized outcome Y , we define

$$Y^+ = -Y - \min(-Y_{min}) + 10^{-6}$$

where Y_{min} is the worst performance in the data and we fix it at 10^{-6} , i.e., the lower bound of the performance range. We adopt this transformation because (i) each phase-level measure (start, riding) and their sum have a meaningful interpretation as *total output*, and (ii) the lowest observed performance is normalized to 10^{-6} .

Team-level. In Panel (b) of Table 2, we summarize 160 team runs (first and second attempts), reporting residualized team finish, start, and riding performance for each attempt, together with starting numbers to account for sequencing effects. By construction, residual means are near zero prior to the transformation; dispersion remains non-trivial after the positive shift. Riding time dominates overall time and exhibits considerable variation, with team riding times ranging from 0 to over 9.09 seconds. These statistics underscore the importance of both individual ability and team-specific complementarities in shaping observed performance.

Panel (c): Nationality. In Panel (c) of Table 2, we tabulate the number of players and team plays by nationality. The sample includes 45 athletes from 12 countries, with substantial variation in team run counts—for example, Canada (CAN) contributes 10 athletes and 62 team runs, while other countries such as Great Britain (GBR) and Slovakia (SVK) appear only once or twice. The relatively limited sample size reflects our focus on athletes with recorded solo performances in monobob events, which allows for consistent identification of individual skill across contexts. Importantly, nationality serves not only as descriptive metadata but also as a realistic constraint on team composition and formation, much like organizational or institutional constraints in firms or federations.

In Figure 5, Panel (a) (Start Time) is tightly clustered and roughly unimodal; terciles based on residualized solo starts overlap heavily, showing limited dispersion in start skill consistent with a standardized, explosive task. Panel (b) (Riding Time) is much wider, with separated terciles, a pronounced upper tail, and a mild shoulder suggesting a subgroup with strong riding. This dispersion captures heterogeneity in line choice, steering, and stability that later informs the team-production estimates.

A.1.3 Correlation between Start-Phase and Riding-Phase Skills

A natural question is whether start and riding skills are distinct or simply manifestations of overall athleticism. Figure 6 plots residualized start skill against riding skill with linear fits. For the 135 athletes with monobob records (Panel a), the correlation is positive. For the 45 athletes in our two-woman sample (Panel b), the correlation is only moderate. Two implications follow: very high-skill monobob athletes are less likely to appear in two-woman events (perhaps preferring to compete solo), and start strength only partly predicts riding control. The moderate correlation supports treating start and riding as separable dimensions, so heterogeneity in implied efficiency reflects task-specific coordination rather than pure overall ability.

Each phase’s solo-skill fixed effect decomposes into phase-specific skill and a non-specific component (e.g., general athletic ability). Correlated non-specific skill across phases can generate correlated individual fixed effects. Thus, high phase affinity in the main text may reflect general

Table 2: Summary Statistics

(a) Player-level

	N	mean	sd	min	max
Residualized Solo Finish Outcome	45	2.19	0.63	0.00	3.65
Residualized Solo Start Outcome	45	0.62	0.22	0.00	1.13
Residualized Solo Riding Outcome	45	1.90	0.59	0.00	3.62

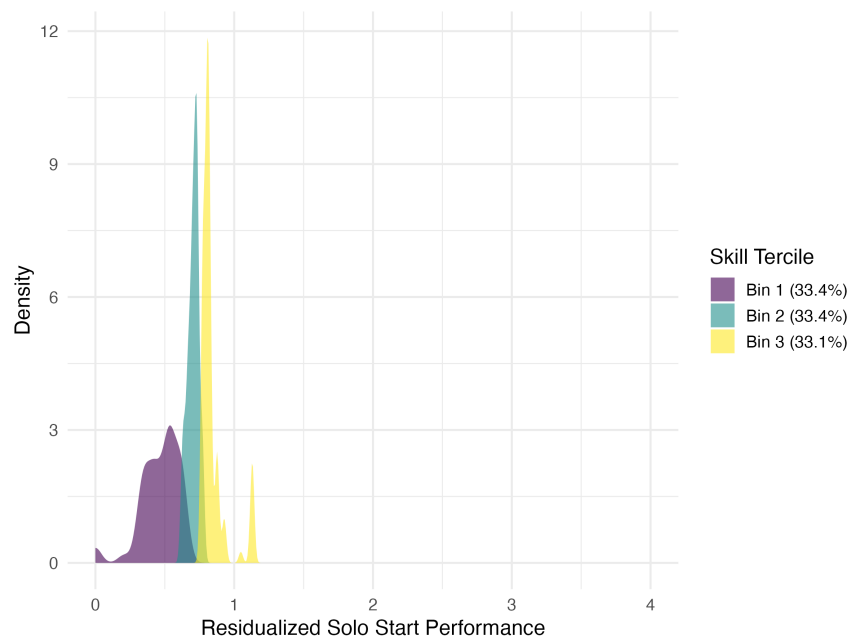
(b) Team-level

	N	mean	sd	min	max
Residualized Team Finish Outcome 1st attempt	160	2.71	0.64	0.00	4.53
Residualized Team Finish Outcome 2nd attempt	160	3.86	0.71	0.00	6.41
Residualized Team Start Outcome 1st attempt	160	0.61	0.13	0.00	0.96
Residualized Team Start Outcome 2nd attempt	160	0.83	0.14	0.00	1.17
Residualized Team Riding Outcome 1st attempt	160	4.90	2.62	0.00	9.09
Residualized Team Riding Outcome 2nd attempt	160	4.55	2.54	0.00	8.81
Starting No	160	7.15	4.14	1.00	19.00

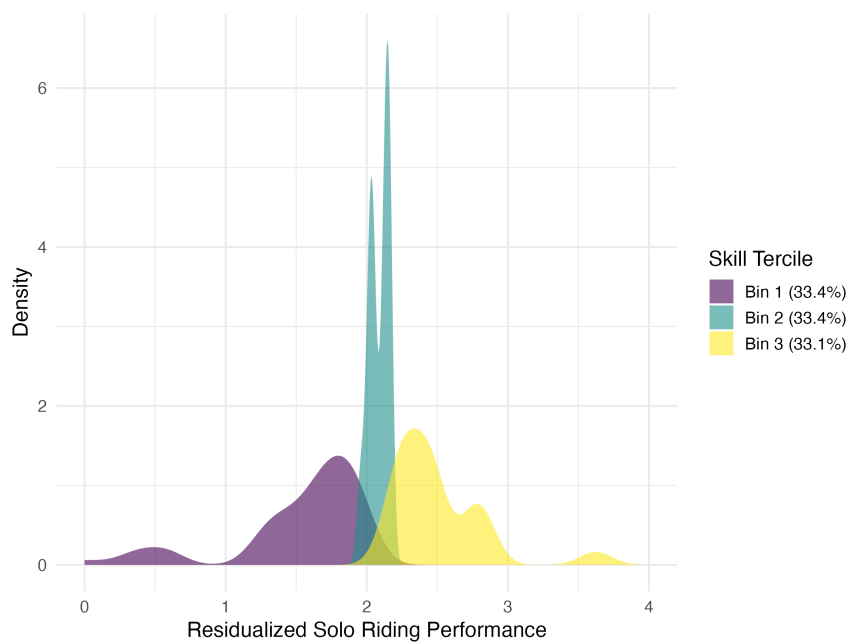
(c) Num of Players and Team Participations by Nationality

Nat	Num of Players	Num of Team Participations
AUS	5	18
AUT	2	4
CAN	10	62
GBR	2	1
GER	5	6
ITA	2	12
KOR	4	3
POL	2	4
ROU	2	19
RUS	3	2
SVK	2	1
USA	6	28

Note: Players in our study are defined as athletes with recorded solo performances in monobob events.



(a) Start Time



(b) Riding Time

Figure 5: Residualized Solo Performance as Individual Skill

non-specific affinity, phase-specific matching, or both.

A.1.4 Descriptive Statistics on Pairing Patterns

This subsection provides descriptive statistics on team pairing patterns in our sample. We examine (i) how many observations fall into each skill-pairing cell when athletes are classified by terciles of their individual skill, and (ii) whether pairing decisions appear to reflect complementary skills across phases (e.g., pairing a strong start-phase athlete with a strong riding-phase athlete) rather than matching on the same skill dimension.

Within-Phase Skill Pairing. Table 3 reports the number of team-run observations in each cell of a 3×3 matrix, where rows correspond to Player 1 (driver) skill terciles and columns correspond to Player 2 (brakeman) skill terciles. Tercile 3 denotes the fastest (highest-skill) athletes, while tercile 1 denotes the slowest.

Table 3: Number of Observations by Skill Pairing

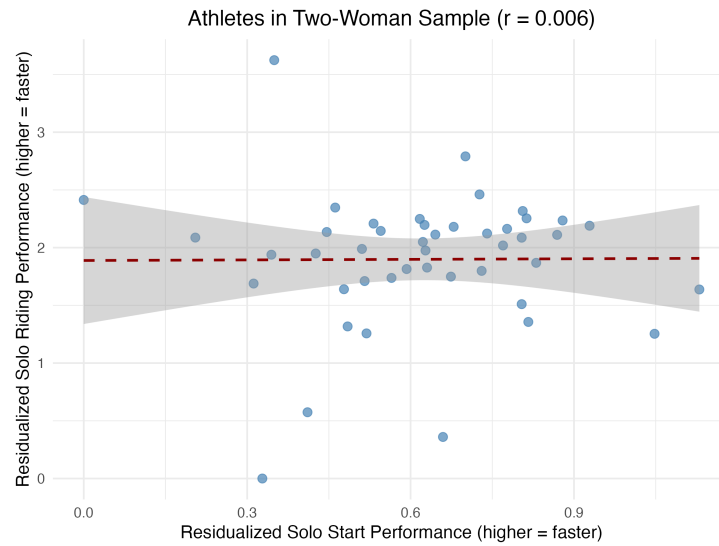
(a) Start Skill			
P1 Start-skill Bin	P2 Start-skill Bin		
	1	2	3
1	31	7	15
2	7	19	27
3	15	27	12
(b) Riding Skill			
P1 Riding-skill Bin	P2 Riding-skill Bin		
	1	2	3
1	6	33	14
2	33	16	4
3	14	4	36

Note: Each cell reports the number of team runs where the driver (Player 1) and brakeman (Player 2) fall into the corresponding start-skill and riding-skill terciles. Tercile 1 = lowest performance; tercile 3 = highest performance.

The pairing patterns reveal several notable features. First, there is substantial variation in cell counts, indicating that team formation is not uniformly distributed across skill combinations. Second, diagonal cells (same-skill pairings such as 1–1, 2–2, 3–3) tend to have relatively high counts in both tables, suggesting positive assortative matching on skill within each dimension. However, off-diagonal cells also contain meaningful numbers of observations, indicating that many teams



(a) Players who participated in monobob races



(b) Players who participated in monobob and two-woman races

Figure 6: Correlation between Start-Phase and Riding-Phase Individual Skills

Note: Each point represents an athlete in the two-woman analysis sample. Skills are measured as residualized solo performance from monobob events, controlling for event and starting order fixed effects, and transformed so that higher values indicate better performance. The dashed line shows the linear fit.

combine athletes with different skill levels. This heterogeneity in pairing provides the variation needed to identify how team performance depends on skill composition.

Cross-Phase Skill Pairing. A natural question is whether coaches form teams based on complementary skills across phases—for example, pairing a driver with strong start skills with a brakeman who excels in riding, or vice versa. Table 4 examines this by cross-tabulating Player 1’s skill in one dimension against Player 2’s skill in the other dimension.

Table 4: Cross-Skill Pairing

(a) Player 1 Start Skill vs. Player 2 Riding Skill

P1 Start-skill Bin	P2 Riding-skill Bin		
	1	2	3
1	18	19	16
2	11	8	34
3	24	26	4

(b) Player 1 Riding Skill vs. Player 2 Start Skill

P1 Riding-skill bin	P2 Start-skill bin		
	1	2	3
1	29	19	5
2	17	13	23
3	7	21	26

Note: Each cell reports the number of team runs where Player 1 (driver) falls into the corresponding skill tercile and Player 2 (brakeman) falls into the corresponding skill tercile.

Table 4 shows dispersed counts rather than off-diagonal concentration, indicating that complementary matching is not a dominant strategy; pairings are broadly positive-assortative on overall ability. Consequently, heterogeneity in implied efficiency likely reflects true affinity rather than strategic matching, consistent with formation driven by constraints such as nationality, availability, or schedules.

A.2 Estimation

Latent efficiencies We estimate latent efficiencies and production functions using a nonparametric strategy.⁸ For a given task k , the observed output $H_{\ell kt}$ is modeled as:

$$H_{\ell kt} = m_k(A_{\ell kt}X_{\ell kt}, Y_{\ell kt}),$$

where $X_{\ell kt}$ and $Y_{\ell kt}$ are team-level skill inputs, and $A_{\ell kt}$ is a latent efficiency term specific to team ℓ in task k and period t . We assume m_k is strictly increasing, continuously differentiable, and satisfies CRS.

Let $G_k(H|X, Y)$ denote the conditional distribution of output for task k given observed inputs. For identification, we exploit the fact that CRS implies:

$$\frac{H_{\ell kt}}{X_{\ell kt}} = m_k\left(A_{\ell kt}, \frac{Y_{\ell kt}}{X_{\ell kt}}\right),$$

which allows us to identify $A_{\ell kt}$ up to scale by holding Y/X fixed and examining variation in H/X .

We estimate $G_k(H|X, Y)$ using kernel-weighted empirical distributions. For a given point (H_0, X_0, Y_0) , the estimator is:

$$\hat{F}_k(\psi A_0 | \lambda X_0) = \sum 1(H_i < \psi H_0) \cdot \kappa(X_i, Y_i; \lambda X_0, Y_0),$$

where $\kappa(\cdot)$ is a kernel function with bandwidth selected to capture local variation. We invert this distribution to recover:

$$A_{\ell kt} = F_k^{-1}(G_k(H_{\ell kt}|X_{\ell kt}, Y_{\ell kt})|X_{\ell kt}).$$

This process is repeated for each task k , yielding estimates of $A_{\ell kt}$ and $m_k(\cdot, \cdot)$ for all tasks. The additive separability across tasks implied by:

$$H_{\ell t} = \sum_{k=1}^K H_{\ell kt} = \sum_{k=1}^K m_k(A_{\ell kt}X_{\ell kt}, Y_{\ell kt})$$

ensures that task-specific estimation can proceed independently, conditional on observed team-level variation.

Elasticity Estimation Finally, to compute production elasticities, we run a second-order polynomial regression that projects observed team output H on polynomial terms of the effective leader-side input AX , the assistant-side input Y , and their interactions. Specifically, we include

⁸See also [Otani \(2024\)](#) for Monte Carlo simulations.

both original and squared terms to allow for flexible curvature in the production function. This estimation yields approximations of the partial derivatives of the production function $m(AX, Y)$ with respect to its arguments, and thus permits calculation of the associated elasticities.⁹

In applied settings, changes in $X_{\ell kt}$ reflect observable training investment, while variation in $A_{\ell kt}$ captures unobserved dimensions such as coordination ability, psychological readiness, or other team-level synergies. Our approach disentangles these two sources of performance and allows for evaluation of training returns and performance diagnostics at the task level.

A.3 Additional Results: Performance Elasticity with respect to Individual Skill Investment

Figure 7 illustrates the performance elasticity with respect to individual performance inputs—capturing the percentage change in team finish performance resulting from a percentage reduction in an individual’s residual start or riding skill—for both players across different performance components and attempts. A positive elasticity indicates that increasing an athlete’s own start or riding skill yields an increase in team finish time performance, reflecting improved performance. Across panels (a) and (b), we observe substantial heterogeneity in the elasticities of Player 1 (the driver) in both start time and riding time. While many drivers exhibit positive elasticities, indicating that their individual time is a meaningful lever for improving team performance, a small share also shows near-zero or negative elasticities, suggesting limited or even counterproductive effects. Importantly, although the magnitude of elasticities varies across individuals, the ranking of drivers is broadly consistent across the two phases: those with high elasticity in start time also tend to exhibit high elasticity in riding time. This implies that while not all drivers contribute equally to performance improvement, those with high responsiveness in one dimension tend to have it in the other as well. Rather than reflecting skill complementarity per se, this pattern indicates correlated effectiveness in increasing individual skill across task components.

In contrast, the patterns for Player 2 (the brakeman) in panels (c) and (d) reveal a more heterogeneous and weakly coupled profile in implied efficiency. In the start phase, only a minority display clear efficiency gains, whereas most observations cluster around zero or indicate efficiency losses, implying that the brakeman’s marginal influence on team performance is limited or even adverse once the driver’s skill is held fixed. In the riding phase, the distribution of implied effi-

⁹The production elasticity with respect to the leader-side input X is defined as: $\frac{d \log m(AX, Y)}{d \log X} = \frac{dm(AX, Y)}{dX} \cdot \frac{X}{H} = \frac{\partial m}{\partial(AX)} \cdot \frac{d(AX)}{dX} \cdot \frac{X}{H} = \frac{\partial m}{\partial(AX)} \cdot A \cdot \frac{X}{H} = \frac{d \log m(AX, Y)}{d \log(AX)}$. To estimate this, we approximate $m(AX, Y)$ with a second-order polynomial: $m = \beta_1(AX) + \beta_2(AX)Y + \beta_3Y + \beta_4(AX)^2 + \beta_5Y^2$, and compute the partial derivatives $\frac{\partial m}{\partial(AX)} = \beta_1 + \beta_2Y + 2\beta_4(AX)$, $\frac{d \log m}{d \log X} = \frac{d \log m}{d \log(AX)} = (\beta_1 + \beta_2Y + 2\beta_4(AX)) \cdot \frac{X}{H}$, and $\frac{d \log m}{d \log Y} = (\beta_2(AX) + \beta_3 + 2\beta_5Y) \cdot \frac{Y}{H}$. These elasticities can be evaluated at observed values or at counterfactual allocations such as the planner-optimal levels X^* and Y^* , and are used in the calculation of the mismatch index based on marginal returns.

ciency is similarly thin and largely orthogonal to that of the start phase; pockets of gains appear idiosyncratic and often differ across attempts. These features are consistent with the brakeman’s operational role—primarily weight placement and stabilization during the run—implying that systematic efficiency improvements attributable to Player 2 are modest and not tightly aligned across input dimensions.

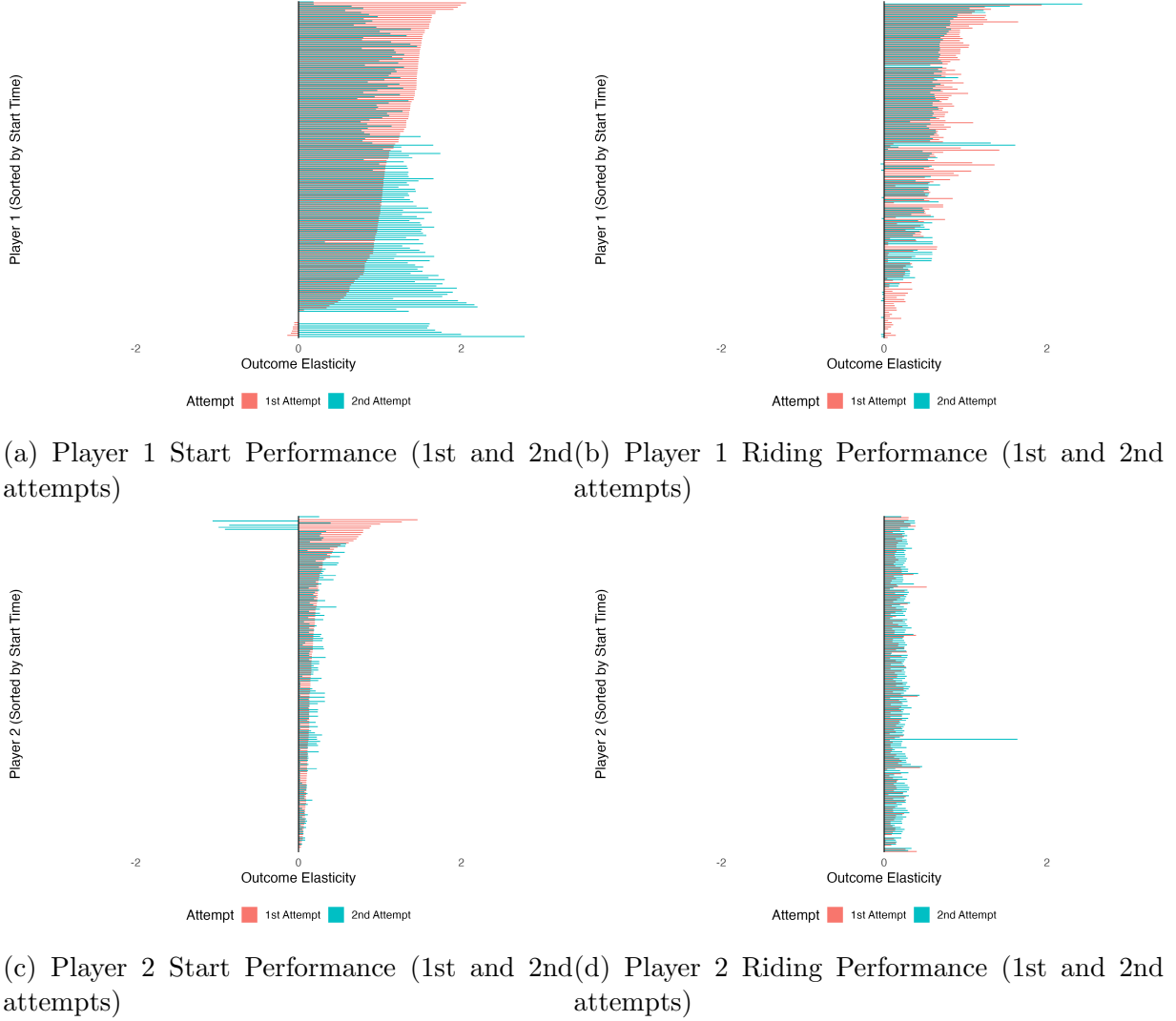


Figure 7: Performance Elasticity with Respect to Individual Skill Investment of Player 1 (driver) and 2 (brakeman)