RIETI Discussion Paper Series 24-E-035

How Much of Merit is Due to Luck? Evidence on the Butterfly Effect of luck

KOIZUMI, Hideto
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-035
March 2024

How Much of Merit Is Due to Luck? Evidence on the Butterfly Effect of Luck*

Hideto KOIZUMI
Research Institute of Economy, Trade and Industry
Hitotsubashi University

Abstract

Progressive societies aspire to eliminate discrimination and promote equal opportunities and meritocracy. The crucial question remains: does the pursuit of equal opportunities and merit-based evaluation truly lead to a fair society? To test the presumed fairness of meritocracy, this paper quantitatively analyzes the impact of luck on merit in the absence of initial differences in individual characteristics. This study utilizes a distinctive experimental setting involving Japanese speedboat races. Participants are randomly assigned engines with different capacities in each tournament, ensuring probabilistic fairness across racers but introducing variability in the timing of luck. By identifying racers who are assigned "lucky" motors during their debut periods as the treatment group, we trace their performance trajectory, examining factors such as the number of first-place finishes and earnings. The findings indicate a growing performance gap over time, accompanied by increased opportunities and tendency for risk-taking behavior. Over four years, the initially modest advantage of the treatment group results in a remarkable 69% more cumulative first-place finishes and 48% more cumulative earnings for male racers. Additionally, male racers in the treatment group exhibit a 10% lower market exit rate compared to the control group. These results underscore the pivotal role of early-stage luck in shaping substantial differences in merit, challenging the presumed fairness of meritocracy.

Keywords: feedback loop, butterfly effect, speedboat race, meritocracy, inequality
JEL classification: D63, D83

*This study is conducted as a part of research at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at the DP seminar for this paper at the Research Institute of Economy, Trade and Industry (RIETI). I would like to thank participants of the RIETI DP Seminar for their helpful comments.

1 Introduction

Progressive societies strive to eradicate discrimination and champion equal opportunities and meritocracy. The pivotal inquiry lingers: does the pursuit of equal opportunities and merit-based assessment genuinely culminate in a fair society? To test the presumed fairness of meritocracy, this paper quantitatively analyzes the impact of early-stage luck in the absence of initial differences in individual characteristics. Illustratively, even minor initial disparities can burgeon into significant distinctions over time, driven by a positive feedback loop—a phenomenon commonly known as the Butterfly Effect. This dynamic is anticipated across various economic spheres, encompassing academic accomplishments among students and the career trajectories of entrepreneurs and CEOs. Enhanced performance not only attracts additional opportunities but also nurtures heightened confidence, fostering greater motivation and diligence, consequently leading to increased success. If these initial disparities arise from chance, it challenges the presumed fairness of meritocracy. To address the question, this research endeavors to conduct a quantitative analysis assessing the extent to which merit is influenced by early-stage luck when individuals, beginning with identical characteristics, including family backgrounds and abilities, engage in competition within a purely meritocratic evaluation framework.

Empirically addressing this question is, however, challenging for two reasons. First, finding a society that operates in the specific way described is difficult. Furthermore, even if such a society exists, estimating the dynamic nature of the Butterfly Effect poses a unique challenge in the social sciences due to the limited availability of long-term, high-frequency performance data in a natural experimental setting. To overcome these challenges, this research leverages a distinctive experimental context involving Japanese speedboat races. In this setting, engines with different capacities are randomly assigned to participants in each tournament. This random allocation ensures fairness among racers on a probabilistic basis but introduces a temporal variation in the actual occurrence of luck.

The findings of this paper reveal a widening performance gap between the treatment and control groups, accompanied by surprisingly substantial Butterfly Effects. Within the first four years since debut, the modest initial shock fueled by early-stage luck accumulates for male racers, resulting in 69% more first-place finishes and 48% more earnings, together with 10% lower market exit rates compared to the control group.

Interestingly, no discernible effects on female racers are observed. Examining the temporary or permanent exit of racers from the market, I find that the treatment induces male racers to remain while prompting female racers to exit. This divergence may be attributed to factors such as marriage or maternity leave, although the underlying mechanism remains ambiguous.

Finally, I provide evidence on the distributional effects. One might expect that the Butterfly Effects are concentrated the most in the top quantiles of the outcome distributions. Surprisingly, the largest proportion of differences driven by early-stage luck is observed within the middle-range group.

To derive these results, I construct a dataset of Japanese speedboat racing, consisting of 6,952,990 race-by-racer observations involving 2,717 racers spanning the period from 1996 to 2022. Among these racers, 1,120 debuted on or after November 1, 2002, for which I have comprehensive data on racer demographics and pre-debut performance, enabling a balance test. The primary analysis focuses on the period from 2002 to 2011, as this timeframe predates significant institutional changes in engine regulations occurring after 2012 and 2014. The resulting sample comprises 464 racers who debuted during this specified period, with an average experience of approximately four years since their debut.

In Japanese speedboat racing, racers are randomly assigned engines with varying capacities, including those deemed as "lucky." The engines play a crucial role in determining race outcomes, influencing the odds of races. Information about the engines is made public before races and is included in race programs. Note that these engines are replaced annually at each speedboat racing stadium.

To identify "lucky" engines, I first compute the winning percentage for finishing within the top two places for every engine at each stadium.[1] Then, for each new racer in the debut term, I tally the instances of being assigned "lucky" engines over the initial five tournaments. To designate roughly half of the racers into the treatment group, a cut-off of the 85th percentile for engines is chosen to qualify as "lucky."[2]

With the treatment, I perform simple regression analyses to examine dynamics and cumulative effects. To alleviate identification concerns, I conduct a balance test of racer demographics and pre-debut performance. I ensure the balancedness between the treatment and control groups, except for racer sex due to a limited sample size of female racers. In response to the imbalance in racer sex, separate analyses are conducted for male and female racers, with a particular emphasis on male racers, given the insufficient power for statistical analyses arising from the small sample size of female racers.

The implications drawn from the results are threefold. First, the immediate policy implication underscores the importance of early-stage intervention. These findings contribute to existing scholarship by elucidating the dynamics of how initial differences in the early stages significantly impact later stages, whether applied to individuals or businesses. Secondly, the results suggest that whether it pertains to personal income tax or corporate tax, a morally justified stance supports a more progressive tax system. Notably, the pronounced Butterfly Effects of early-stage luck are identified even when the initial characteristics are *identical* between the treatment and control groups. This undermines the basis for meritocratic arguments against progressive tax systems, as a substantial portion of merit appears to be attributed not to innate ability or *voluntary* efforts but rather to sheer luck. Early-stage luck results in more opportunities, subsequently leading to increased efforts, heightened confidence, and ultimately

[1] Winning percentages for finishing within the top two places are used by the organizer to notify the performance of engines for racers and bettors.

[2] The main results are mostly unchanged with different cut-offs of the 80th percentile and the 90th percentile. I will come back to this at the data section.

elevated ability over time.

Last but not least implication is that equal opportunities with a meritocratic evaluation system will not lead to a fair society with justifiable inequality. The results of this paper unequivocally demonstrate the significance of the role of early-stage luck. It is important to note that the luck studied in this paper is posterior, occurring after birth, rather than the prior luck determined by genes and family backgrounds. This implies that eliminating disparities caused by prior luck does not necessarily eradicate discrepancies induced by everyday luck. Therefore, it becomes challenging to claim that we *deserve* our current outcomes, as these outcomes can be largely influenced by chance. Michael Sandel argues that the prevalent meritocratic claim in contemporary societies significantly diminishes the self-esteem of many non-elites, fueling their resentment towards elites and contributing to recent populist movements worldwide (Sandel (2020)). The findings of this study challenge the validity of the meritocratic claim, providing some empirical support for the *redistribution of esteem* as Sandel (2020) proposes.

This paper contributes to multiple strands of the literature. It adds to the established literature on the enduring effects of early differences in abilities and life conditions. Within the literature that delineates the so-called "relative age" effects of early childhood maturity differences induced by the school start dates and variations in birthdays, closely related studies demonstrate the lasting impact on various outcomes including university enrollment (Bedard and Dhuey (2006)), labor earnings (Kawaguchi (2011)), top management (Du et al. (2012)), and political positions (Muller and Page (2016)).[3] Notice the presence of an ability gap between early-born and late-born children within a class in this context. Another strand of the same literature proves the importance of early life conditions including genes in shaping multiple life skills that culminate in sizable differences in socioeconomic outcomes later in life (see, e.g., Cunha et al. (2005); Heckman and Mosso (2014); Bailey et al. (2021); and Houmark et al. (2024)). This paper complements the literature by (i) providing evidence on the enduring effects of early-stage luck even if ability and other life conditions such as family backgrounds are the same to start with, (ii) demonstrating that adults also experience persistent and escalating effects of small advantages early in their careers, and (iii) portray *how* the gap evolves over time in a highly controlled environment that excludes factors like family backgrounds, neighborhood effects, and nepotism that interact with the evolution of childhood performance in the literature.

Finally, this literature contributes to a strand of papers that find the long-term effects of external factors on performance in general. Almond (2006) investigates the long-term effects of the Spanish flu on economic and socioeconomic outcomes, while Maccini and Yang (2009) examines the impact of childhood rainfall on later-life socioeconomic outcomes. Additionally, Genda et al. (2010) explores the effects of economic downturns on subsequent outcomes in labor markets. In contrast, my paper contributes unique evidence on the dynamic and cumulative effects of early-stage luck on actual *ability*,

[3] Yamaguchi et al. (2023) provide empirical evidence on non-cognitive skill formation differences induced by the relative age effects, potentially serving as a mechanism behind some of these long-term effects.

a dimension not measured by these studies.

Ginsburgh and Van Ours (2003) exploits a setting that is more akin to this paper than these papers. They show the effects of the random ordering of finalists at a world-renowned musical competition on the later-stage market success of the participants. They measure success by the presence of the historical participants' names on the catalog of a Belgian library together with British and French catalogs of classic music players. It is important to note that they were interested in the effects of expert opinions on market recognition and thus market price, unlike this paper that investigates the effects of luck on merit including actual *ability* and earnings. This paper adds to the literature by providing (i) the intermediate dynamics, (ii) analysis on the population of new competitors, and (iii) direct measures for merit. The high-frequency nature of my dataset allows me to delve into the dynamics of luck. Furthermore, unlike the finalists who are at the very top of young pianists, my setting covers the population of the competitors in a sport, allowing for an estimation of distributional effects. Lastly, although their catalog indicator is a great measure for success, it may not perfectly capture ability since, as the focus of their paper, expert opinions may affect the market prices of the musicians through the demand side. They also write, "[e]arnings would, of course, have been a much better choice, but are impossible to collect" (p. 290). The number of first-place finishes and earnings in a sport allows me to directly measure merit, avoiding the influence of expert opinion.

2 Speedboat Racing in Japan

For the description of speedboat racing in Japan, I reference Booth and Yamamura (2018) for certain aspects. Note that for the purpose of this study, I augment this information with greater details on the engine allocation process, engine structures, and historical changes in institutional regulations. Additionally, I provide fewer details on gender differences, as my study addresses a distinct question than theirs. For the additional details, I rely on Himura (2015), the principal source in Booth and Yamamura (2018), as well as sources such as Fujino (2006), Himura (2017), the official website of the Japanese Speedboat Racing Association (JSRA), and various online articles, including those authored by Himura.

Speedboat racing in Japan is organized in the form of tightly regulated tournaments overseen by the Japanese Speedboat Racing Association. Both male and female racers undergo identical intensive training at the single training institution, the Yamato Boat School.[4] Individuals aged between 15 and 29 aspiring to become professional speedboat racers are required to undergo a year of training at the Yamato Boat School and pass a final examination. There are two timings in a year when they graduate from this school and become professional racers: May 1 and November 1. The wide age range allows individuals from diverse backgrounds to enter the profession, ranging from those with junior

[4]Women not only train alongside men in similar conditions but also engage and compete in races under the same circumstances.

high school education to university graduates, some of whom have pursued other careers after their education. Therefore, the time elapsed since graduating from the boat school varies among entrants.

According to Himura (2015), there are approximately 1,600 speedboat racers in Japan, spanning ages 18 to 70, with around 1,400 men and 200 women. While the number of female racers has been increasing due to increasing demand for female racers, in my sample period, roughly speaking, 25 % of tournaments are mixed-sex, 5% are female-only, and 70% are male-only. Racers are randomly assigned to mixed-sex or single-sex tournaments. The country hosts 24 speedboat racing stadiums, with races taking place about four days a week in each stadium. There are two terms for evaluation in a year: the first term ranges from May 1 to October 31, while the second term ranges from November 1 to April 30.[5] The racing circuit consists of a 600-meter artificial pond or sectioned-off water body, around which competitors race three times, totaling a distance of 1,800 meters.

As depicted in Figure 1, speedboat racing employs the premature start system, requiring boats to cross the starting line within one second after the starting clock hits zero. Initially, racers are assigned their pits by the association committee. Nevertheless, racers have the strategic option to alter their starting lanes during the pit-out period as described in Figure 1, potentially resulting in a different starting position for the race. Being positioned directly behind another boat is considered a violation. In general, for safety reasons, new racers are assigned the most outward pits (6th lane). This makes it hard for new racers to win a race unless they are assigned a good engine.

2.1 Tournament Grade, Racer's Grade, Prizes, and Penalties

Tournaments are categorized into either graded or non-graded (so-called "usual races") tournaments. Graded tournaments are further classified into four sub-classes, ranging from the lowest rank grade III (GIII) to grade II (GII), grade I (GI), and Special Grade (SG). In general, non-graded, "usual race" tournaments are bottom-graded and SG tournaments are the top-graded. Each tournament is held at a stadium over three to seven days (usually four to six days) with a racing meet on each day. Each racing meet comprises twelve races, each with six racers competing. These races progress through various stages, from preliminaries and semi-finals to the finals.

A racer earns points based on their finishing positions in races.In non-graded and GIII tournaments, points are awarded as follows: 10 for first place, 8 for second, 6 for third, 4 for fourth, 2 for fifth, and 1 for sixth. GI and GII tournaments add 1 point to these values(Himura (2015)), while SG tournaments add 3 points prior to the first term of 2005 and 2 points thereafter.[6] Penalties, such as those for poor navigation or rule violations, can result in a loss of 7 points.

Consider a GIII tournament spanning six days. Racers participate in multiple preliminaries from

[5] The JSRA applies the results of evaluation from the first term to January 1 till June 30 next year, and those of the second term to July 1 till December 31.

[6] From https://sp.macour.jp/columns/himura/154720/ accessed on November 2, 2023.

day 1 to day 4, with the top 18 moving on to semi-finals. Racers not making the cut participate in “standard competition” races on day 5 and 6. Most of the tournaments employ the three semi-finals system, while few employ other systems such as two semi-finals and no semi-final. Generally speaking, those racers with higher points will be assigned more inward and thus better initial pits and lanes than the other racers on day 5. Day 6 features finals, “specially selected A group” races for the third and fourth-place racers, “specially selected B group” races for the fifth and sixth-place racers, and another round of “standard competition” races. The number of races within a tournament remains the same for racers and engines in general.

Not only tournaments are graded, but also racers. There are four racer grades: A1 (the top racer grade, around 20% of all racers), A2 (around 20% of all racers), B1 (around 50% of all racers), and B2 (the bottom racer grade, around 10% of all racers). Individual racers’ grades change twice a year (the first period is from January 1 to June 30, while the second period is from July 1 to December 31) and the change is based on the number of points and races in the most recent term.[7] For example, for a racer grade from January 1 to June 30, 2023, the performance during the first term of 2022 (May 1 to October 31, 2022) is used. To be graded as an A1 racer, a racer has to achieve the following three criteria. One is the winning percentage for finishing within the top two places higher than 30% and the winning percentage for finishing within the top three places higher than 40%. Another is a boat accident/penalty rate below 0.7, which will be explained shortly. The last criterion is the number of races attended. This last criterion was modified from 70 races to 90 races and became effective as of the second term of 2010.[8] Among those who satisfied these criteria, some top racers are awarded A1 grade, so that the top 20% of all racers are graded A1. These A1 racers are eligible for most of the graded tournaments, and among these A1 racers, some will be selected to participate in SG tournaments. A2 grade has the same criteria, except for the minimum number of races to remain at 70 even after the second term of 2010. Those racers who satisfy these criteria but are not screened into A1 grade will be awarded A2 grade in a way that the proportion of A2 racers becomes roughly 20%. A2 grade racers are also eligible for most of the graded tournaments except SG tournaments, while an A2 racer is eligible for an SG tournament if he or she is the winner of the SG tournament in the previous year. The vast majority of non-A1 and non-A2 racers will be categorized into B1 grade, except new racers and those racers who caused boat accidents in general. The criteria for B1 grade changed in the first term of 2016. Prior to this term, the criteria were (a) the winning percentage for finishing within the top two places above 10%, (b) the winning percentage for finishing within the top three places above 20%, and (c) the minimum number of races above 50, so that 50% of all racers are classified into B1. These criteria, however, became too strict to achieve 50% proportion toward 2016. The new criteria are criterion (c) plus (d) total points

[7] Booth and Yamamura (2018) write that these grades are determined based on points aggregated for three years, but it must have been a typography for a half year. Note that each racer’s license renewal is conducted ever three years.

[8] From http://www.kenkyu.co.jp/station/archives/2010_10_4_106.html accessed on November 2, 2023.

in the previous term divided by the total number of races in the previous term is above 2, without criteria (a) and (b).[9] To prevent racers from earnings points by attending lower-graded or non-graded tournaments only, the Association assigns racers to tournaments to equalize opportunities as much as possible for non-graded tournaments. In general, racer grade A1 and A2 racers are allocated races of approximately 15 days, B1 around 12 days, and B2 around 8 days within each month, and each non-graded tournament consists of racers proportional to the distribution of these racer grades. A racer can decline tournament participation offers from the JSRA for reasons like family issues, but a racer cannot decline an offer to attend one tournament and choose to participate in another tournament that takes place during the same period.

The prizes offered are substantial but vary significantly across tournaments and the stages of tournaments explained above. The amount of prizes on average is ordered from SG to G3 tournaments in a descending manner. In general, graded tournaments have higher average prizes than non-graded tournaments. The amounts of SG tournament prizes range from around 18 million to 110 million yen depending on tournaments, while those of non-graded tournaments tend to be slightly more than 0.74 million yen. Notice that the discrepancies of prizes between tournaments are much larger than those of the aforementioned points. Therefore, points are far from precise measures of earnings.

There are two components in the prize amounts. One is compensation for participating in races. Then, even the sixth (last) place in each race offers some money to participating racers. The other part is the actual prize/award for placements. Note that every placement generally offers some prizes that are the highest for the first place and lowest for the sixth place. The amounts of these prizes are higher in the semi-finals and even higher in the finals than the standard competition races within a tournament.

2.2 Engines

On a day before a tournament, boats and engines are randomly allocated to racers participating in the tournament.[10] The same allocated boats and engines are used throughout the tournament. For random allocation, lotteries are publicly conducted. In general, each stadium provides two separate lottery wheels for boats and engines (one for boats, and the other for engines). Each racer draws a ball from each lottery wheel in front of the other racers. Racers are not permitted to refuse the boat or engine randomly assigned to them. These boats and engines are made by the same company for all stadiums and are annually renewed, but the renewal month of a year varies across stadiums. As Himura (2017) explains, boats do not affect the results of races as much as engines do. He recommends not to care about the past performance of boats such as the winning percentage for finishing within the top two

[9] From https://www.boatrace.jp/owpc/pc/site/news/2016/03/1236/ accessed on November 2, 2023.

[10] See an actual lottery for motor and boat allocation at https://www.youtube.com/watch?v=r1OTsnCeJGc accessed on November 2, 2023.

places unless it is terribly poor, while he emphasizes the significance of engines.[11]

This is because of a reason caused by an assembly of many parts similar to horsepower differences among automobiles of the same brand due to individual differences among the capacities of the same engines. Speedboat engines (motors) are made of around 400 different parts. While the production process is standardized, some margins of errors are permitted. When rotational speeds of motors attain high that causes heat, these small errors get amplified, due to so-called engine "bloating." This will lead to slower piston moves in the engine and thus shaft. According to Himura (2022), there are significant capacity differences among new engines even in the very first race of these engines. Racing programs provide data on the performance of each engine in the past races such as the winning percentage for finishing within the top two places, and bettors use this information.

Racers are allowed to fine-tune engines, which requires some experiences. While there are around 400 parts in an engine, a racer is allowed to replace only nine kinds of them: piston, ring, carburetor, gear box, electrical system parts, crankshaft, cylinder box, screw propeller, and carrier body. To replace these parts, they have to consult with a mechanic and obtain his or her permission. The replacement costs are free in general.

Before April 1, 2012, racers were allowed to bring their own propellers. This system was abolished mainly because racers with enough capital could pay external companies to make high-performance propellers, which made the betting hard.[12] After 2012 April 1, racers have to use propellers that accompany assigned engines. They are allowed to tune up the propellers using wooden hammers to adjust the propellers. These tuning skills take some experience and make it difficult for new racers to adjust them for their racing styles.

This system change makes it hard to measure the capacity of an engine at one point in time since what is "good" depends on whether or not the previous user happened to have adjusted the propellers in a way that it matches the racer's style, especially for racers with less experience. Section 3.2 empirically prove this. Since I cannot tell if the adjustment done by the previous user matches a particular new racer's racing style, I focus on the period before 2012 and those racers who debuted before October 31, 2011.

Note that the JSRA introduced a new model of engine (Yamato 331) in December 2014. The new model has much less horsepower,[13] and this makes it even harder for new racers to win a race since they are assigned the most outward pit and lane. I empirically show this in section 3.2. By the end of 2015, all stadiums replaced entire engines with the new model. Additionally, the new model introduces

[11] Bettors usually ignore boat performance but carefully check engine performance. See https://wsobv.com/boatrace/7567 and https://kyoutei-navi.com/beginner/motor/ accessed on November 2, 2023.

[12] Speedboat racing society adopts the apprenticeship system, and mentors sometimes give mentees propellers. The random assignment must equalize the probability of having a mentor with good propellers.

[13] See an article about this at the JSRA website, https://www.boatrace.jp/owpc/pc/extra/race/24_20150526/monthly/himuken.html.

a larger variance in capacity than the old model.[14]

Right after random draws of boats and engines, racers are provided with racing programs for the most recent past tournaments that have information on the winning percentages for finishing within the top two places of engines and which racers were assigned these engines. Given the information, racers determine strategies during the actual races.

2.3 Strategies: lane changes

As explained above, while there are initial lanes assigned to racers, racers are allowed to change lanes during the premature start. However, while lane changing can bring benefits, it can also bring costs for two reasons. One comes from the rule that prohibits racers to stop their engines and break before starts. Lane changes toward more inward lanes require racers to accelerate more than the other racers, and this makes it more likely for racers to position too close to the starting line. This results in insufficient approach distances, which leads to insufficient acceleration before starts. The second reason comes from rule violation risks. The race rules are strict, and breaking them leads to serious penalties. Therefore, lane changing can be considered a risk-taking behavior and thus is used as a measurement of risk-taking as used by Booth and Yamamura (2018). The details of the penalties are summarized in the next section.

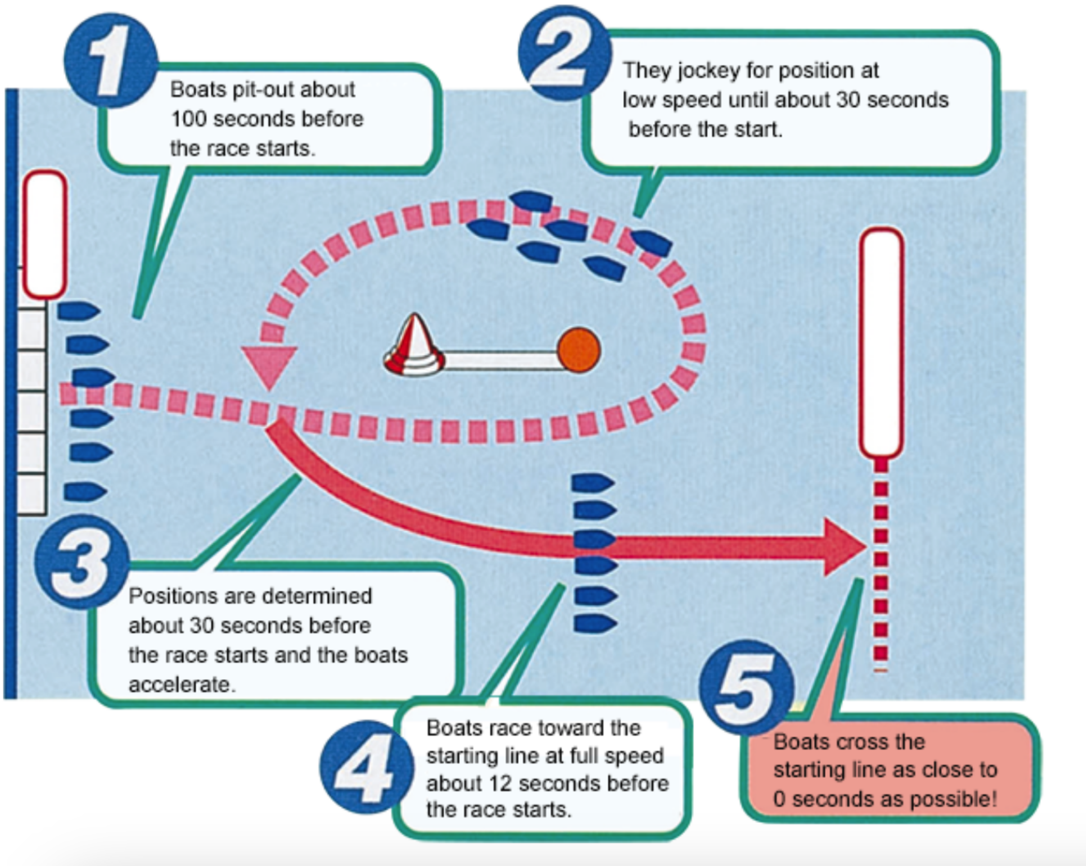
2.4 Penalties and Retirement Admonishment

The JSRA sets penalty points for the following acts: (a) flying or late start at the finals of any tournament (30 points), (b) flying or late start at the non-finals of any tournament (20 points), (c) disqualification for interference (15 points), (d) disqualification or absence in a race for a racer's lapse (10 points), and (e) (minor) bad boat navigation and violation of standby behavior (2 points). If a racer's penalty rate (total penalty points divided by the number of races) within one term surpasses 0.7, or if the racer's total number of races attended within one term does not reach 60, this racer will be downgraded to racer grade B2. Furthermore, every flying or late start will result in the cancellation of tournament assignments for 30 days following the last tournament that has already been assigned and scheduled for the penalized racer.

If a racer poorly performs throughout the most recent four consecutive terms, he or she is "recommended" to retire. While the recommendation is not enforceable, most racers do quit following the recommendation. The JSRA sets the standard for recommendation as either (i) the winning rate (the sum of race points explained above divided by the number of races) of the most recent four consecutive terms is lower than 3.80, (ii) penalty rates of the most recent four consecutive terms is above 0.7, or (iii) the total number of races attended throughout the four consecutive terms is below 60. Note that this

[14] See the same article, https://www.boatrace.jp/owpc/pc/extra/race/24_20150526/monthly/himuken.html.

Figure 1: The Premature Start System



Source: Japan Boat Race Association, https://www.boatrace.jp/owpc/pc/extra/en/index.html(accessed November 2, 2023).

rule does not apply to new racers for the first three years. Then, the first retirement recommendation can take place at the end of the fifth year of racers since their debut.

3 Data, Treatment Definition, Stylized Facts

The main dataset comes from publicly available data provided by the JSRA. The institution provides all the race results of all stadiums that took place after July 17, 1996. The race results include the race dates, stadium names, tournament names, race distance, weather, wind strength, wave height, placement (arrival order and any rule violation), initial lane position, racer names, racer ID, race categories (e.g., semi-finals), engine ID, boat ID, exhibition time records for test runs, lanes after lane changing, start time, and race time.

The JSRA also provides demographics of racers after October 31, 2001 that get updated every term. The demographics include racer grades, birth dates, sex, age, height, weight, and blood type. Also, I was able to obtain pre-debut performance of each racer who debuted only on and after November 1, 2002. While I will explain the definition of the treatment in the next section in more detail, the treatment is based on frequencies of being assigned in the debut term engines with relatively good capacities. Therefore, to avoid reverse causality, when I compute engines' first-place percentages, I exclude the race outcomes of those racers who debuted in the term of the race. Given that every motor gets replaced on an annual basis, I do not use observations before November 1, 2001.

This procedure leaves me 6,952,990 race-by-racer observations with 2717 racers. Among these racers, 1120 racers debuted on and after November 1, 2001. As explained above, I focus on the period between November 1, 2002 and October 31, 2011, which leaves 504 racers who debuted in this period. While I will explain the details of the treatment definition in the next section, I will use the first five tournaments to determine the treatment group. Then, I will exclude 28 racers who did not attend five tournaments during their debut term, mostly because they have gotten penalized for flying or late starts within their first four tournaments. I focus on these racers and trace the effect of initial shocks over time. Some of these racers exit from the market. With the variation in the timing of debut and market exits, I use the unbalanced panel of 464 racers.

As for the prize amounts, they are not included in the race results, and earnings are not recorded in the racer demographics. Therefore, I manually collect prize amounts from various sources as much as possible. The details of the data sources are explained in the Appendix. I admit that there will be some measurement errors in the prize amounts (and therefore earnings) coming from both time-series and cross-sectional sources.

The prize amounts of tournaments can fluctuate over time. The prize amounts are affected by the sales of stadiums and the JSRA, while the prize amounts I hand-collected are those of 2023. It is known that the sales of the JSRA steadily decreased from 1991 when the bubble was about to burst to 2010, but they reverted back and have steadily increased since 2010. Given my main sample period, there can be some under- or over-estimation on earnings outcomes. Note that the price level in Japan for the last 30 years is famous to be quite stable.

The cross-sectional measurement error simply comes from incomplete coverage of tournaments. The data sources I use cover many graded tournaments and the lower bounds for non-graded ones. However, they do not cover some graded tournaments and most of non-graded ones. To deal with this, I take the following procedure. If the prize amount of the winner of an entire tournament in the end of March 2023 is known, I use another tournament that has the full list of prize amounts and compute prize amounts for all placements other than the winner using the proportion of winner prize amounts between two tournaments. If the 2023 prize amount for the winner is not known, then I take a conservative approach and use the lower bound of tournaments in the same grade. For non-graded tournaments, I also take a conservative approach and use the lower bound. Therefore, the cross-sectional measurement errors tend to cause a downward bias in my estimates.

The outcomes of interest throughout this paper are (a) the number of first places, (b) earnings, (c) the number of GII or above tournament races, and (d) the proportion of races in which a racer changes his or her initial lane toward the first lane. Together with exit rates, outcome (a) and (b) are meant to capture the merit or success of individual racers, (c) is to measure the quantity of rewarded opportunities for their merit since the eligibility for graded tournaments including racer grades become much stricter for GII or above tournaments compared to GIII, and (d) is to quantify risk-taking behavior.

3.1 Definition of Treatment

The definition of the treatment is as follows. Every year, some new graduates from the Yamato school debut on May 1, while others debut on November 1. I identify which new racers are lucky to be assigned good engines in the beginning of the debut term at least once. Since there are slight variations in the number of tournaments new racers attend, I use the first five tournaments. The reason for choosing five as the cut-off is to balance the sample size between the treatment and control groups to maximize the detective power.[15] Note that average racers are assigned eight tournaments during their debut term.

To define lucky engines, I first compute the winning percentage for finishing within the top two places of every motor at each stadium excluding those new racers outcomes in their debut terms. Then, since the random allocation of engines is at the racer-by-tournament level, for each new racer in the debut term, I count the number of assignment of lucky engines whose winning percentages for finishing within the top two places are at or above the 85th percentile. I chose the 85th percentile so that roughly

[15] Furthermore, increasing it to six tournaments disproportionately excludes racers who were not assigned a lucky engine in the first five and were induced to conduct false starts to achieve better placements. On the other hand, decreasing it to the first four significantly reduces the number of racers assigned lucky engines.

half of the new racers in the study period lie in the treatment group. Changing the percentile cutoff by five percentiles change the proportion of the treatment group by more than 10 percent. The main results remain unchanged with the 80th and 90th percentile cutoffs except that the opportunity measure by the number of attending GII or above races becomes statistically insignificant under the 80th percentile, due to the loss of statistical power.

Table 1 shows the results of the standard balance tests for the sample using the values from or prior to debut terms. In Table 1, the first row is the proportion of male racers, the second row racer age, the third row racer height, the fourth row racer weight, the fifth row racer blood type being type B, the sixth row whether or not racers are born in April or May, and the last row, pre-debut winning rate, corresponds to the pre-debut performance (the sum of race points explained above divided by the number of races) at the Yamato school. At the Yamato school, there are seven tournaments throughout training periods, and those who have high winning rates are specially treated and assigned the number of races equivalent to B1-graded racers.[16] The balancedness of being born in April or May is tested since school starts in April in the Japanese education system, and it is known that early-born children have advantages. Table 1 generally demonstrates the balancedness in each of these variables between the treatment and control group, except for racer sex.

There is no discrimination under the random assignment, and this appears to come from a pure chance due to the small sample size. Then, I conduct balance tests for male and female samples, separately. Table 2 shows a well-balancedness between the treatment and control groups within the male sample. Table 3 displays a balancedness as well, although the height and weight differences are (almost) statistically significantly different between the treatment and control groups within 73 female racers.

These results motivate me to analyze male and female samples separately. When the female sample is analyzed, I include initial heights and weights as controls within regressions. In particular, in a dynamic analysis, I include fixed effects for time interacted with heights and fixed effects for time interacted with weights, while in a cumulative analysis, I simply include heights and weights covariates.

3.2 Engine and race outcomes

Before introducing econometric specifications, I shall check if good engines actually affect the race outcomes of new racers. Using the debut terms of new racers, I conduct an OLS regression of the indicator of the first-place finish on lucky engines with different thresholds at the race-by-racer level. For example, in Table 4, *lucky_60* is an indicator that is equal to 1 if the assigned engine is at or above the 60th percentile of the rate of first and second place finishes. One can see from Table 4 that (a) lucky engines do matter for race outcomes, (b) the higher the performance of the assigned engines, the higher the

[16] The treatment cutoff is three-folded: (i) the winning rate above 6.50, (ii) the winning percentage for finishing within the top two places higher than 30%, and (iii) the penalty rate below 0.7.

Table 1: Balance Test

	(1) Control	(2) Treat	(3) Control - Treat
Male	0.807	0.876	-0.068*
	(0.026)	(0.021)	(0.034)
Age	21.242	21.145	0.097
	(0.182)	(0.138)	(0.226)
Height	163.386	164.116	-0.731
	(0.382)	(0.286)	(0.473)
Weight	52.291	52.568	-0.277
	(0.227)	(0.175)	(0.284)
Blood type B	0.238	0.212	0.026
	(0.029)	(0.026)	(0.039)
Born in Apr or May	0.193	0.220	-0.027
	(0.026)	(0.027)	(0.038)
Pre-debut Winning Rate	5.119	5.295	-0.176
	(0.080)	(0.080)	(0.113)
N	223	241	464

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Balance Test of Male Sample

	(1) 0	(2) 1	(3) (1) vs. (2)
Age	21.478	21.246	0.231
	(0.215)	(0.147)	(0.255)
Height	165.389	165.090	0.299
	(0.275)	(0.249)	(0.370)
Weight	53.372	53.100	0.273
	(0.195)	(0.160)	(0.250)
Blood type B	0.228	0.190	0.038
	(0.031)	(0.027)	(0.041)
Born in Apr or May	0.200	0.227	-0.027
	(0.030)	(0.029)	(0.042)
Pre-debut Winning Rate	5.284	5.419	-0.136
	(0.089)	(0.082)	(0.121)
N	180	211	391

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

Table 3: Balance Test of Female Sample

	(1) 0	(2) 1	(3) (1) vs. (2)
Age	20.256	20.433	-0.178
	(0.218)	(0.383)	(0.413)
Height	155.000	157.267	-2.267*
	(0.754)	(0.676)	(1.065)
Weight	47.767	48.833	-1.066
	(0.363)	(0.437)	(0.567)
Blood type B	0.279	0.367	-0.088
	(0.069)	(0.089)	(0.112)
Born in Apr or May	0.163	0.167	-0.004
	(0.057)	(0.069)	(0.089)
Pre-debut Winning Rate	4.432	4.420	0.012
	(0.144)	(0.215)	(0.249)
N	43	30	73

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

Table 4: Effects of engines on the probability of the first place finish by new racers

	(1) firstplace	(2) firstplace	(3) firstplace	(4) firstplace
lucky_60	0.0074***			
	(0.0019)			
lucky_70		0.0080***		
		(0.0020)		
lucky_80			0.0109***	
			(0.0024)	
lucky_90				0.0160***
				(0.0033)
Control Mean	0.0264	0.0270	0.0273	0.0281
N	32817	32817	32817	32817

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

probability of the first place finish by a new racer, and (c) the probability of the first place finish by a new racer is low in general.

On the other hand, Table 5 shows the same results with the sample of racers who debuted after April, 2012. As explained above, after April, 2012, racers are not allowed to use their own propellers. This means that new racers have to use propellers that were tuned by the racers in the previous tournament. However, new racers have little experience of tuning up propellers by wooden hammers and have difficulties with assigned engines that were tuned up for the use of different racers. This will introduce a large error in the performance of lucky engines for new racers after April, 2012. The standard errors in the estimates of Table 5 support this. Furthermore, as explained above, the introduction of Yamato 331 model engines make it even harder for new racers to win a race. The control means and small estimates of Table 5 evidence this claim.

Table 5: Effects of engines on the probability of the first-place finish by new racers after April 1, 2012

	(1) firstplace	(2) firstplace	(3) firstplace	(4) firstplace
lucky_60	0.0025**			
	(0.0009)			
lucky_70		0.0026**		
		(0.0009)		
lucky_80			0.0027*	
			(0.0011)	
lucky_90				0.0033*
				(0.0014)
Control Mean	0.0056	0.0059	0.0061	0.0063
N	38185	38185	38185	38185

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 Econometric Framework

All the specifications below include racer heights and weights as controls whenever the sample includes female racers, due to (almost) statistically significant unbalancedness in these variables within the female sample. With that said, I employ two specifications for main analyses. One is to trace the dynamics of the Butterfly Effect of luck at an early stage. In this specification, using the debut term as the reference period, I trace the differences between the treatment and control groups in the first eight years by estimating the following equation:

4.1 Dynamic Effects

$$y_{it} = \delta_t + \sum_{t \neq 0} \beta_t D_i \lambda_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is racer i's outcome at term $t \in \{0, 1, ..., 15\}$ (0 corresponds to the debut term, 15 to the end of the racer's eighth year), and D_i is the treatment indicator that equals 0 if racer i is in the control group and 1 otherwise, $\lambda_t = 1$ if event time $= t$, and I include the debut term $t = 0$ as the reference period. The outcome variables include the total number of first places, the natural logarithm of earnings, the total number of GII or above tournament races, and the proportion of races in which racer i changed lanes toward the first lane. Note that each t consists of different sample sizes due to a variation in debut terms and market exits.

For the total number of first places and GII or above tournament races, since there are excessively many zeros especially in early periods, my main analysis for dynamics does neither take the natural logarithm of these outcomes nor use some nonlinear models such as Poisson regression which would cause a bias in the estimates.[17] These outcomes have much fewer zeros when I take the cumulative

[17]Regarding the natural logarithmic transformation, the conventional way of $log(y_{it} + 1)$ would cause a bias in the form derived by Bellégo et al. (2022). As for nonlinear models, Angrist and Pischke (2009) discusses the issues with nonlinear models for limited dependent variables as to biases. If nonlinear regression produce unbiased estimates in a particular setting, then so will a linear regression. However, if the underlying error term does not follow the distribution assumed in the nonlinear models,

values. Therefore, I take the natural logarithm of these variables when I conduct cumulative analyses to take care of the fat-tail skewness in the cumulative outcomes. For the skewness of outcomes that raise concerns about distributional effects, as suggested by Angrist and Pischke (2009), I conduct quantile regressions for these outcomes in the following cumulative analyses.

4.2 Cumulative Effects

The other specification captures the cumulative effects of the treatment, which is as follows:

$$Y_i = \gamma_0 + \gamma_1 D_i + u_i, \tag{2}$$

where Y_i indicates the cumulative version of the above outcomes. For racer i who debuted eight years before October 31, 2011 and who did not exit from the market during the first eight years, Y_i is a eight-year cumulative outcome. On the other hand, for example, if racer i debuted five years before October 31, 2011 without market exits or debuted more than five years before October 31, 2011 but exits from the market in his or her fifth year, then Y_i corresponds to a five-year cumulative outcome. For the proportion of races in which racer i changes his or her initial lane toward the first lane, I take the proportion after taking the cumulative values for both the numerator and denominator.

For both equations (1) and (2), since the treatment is generated by random assignments, the identification threat comes from initial differences between the treatment and control group. The threat is relieved by the results of Table 1.

4.3 Average Market Exit Rate

To analyze whether the treatment affects racers' market exit rates, in addition to the dynamic effects specification above that loses statistical power due to restriction to a pair of specific periods, I also estimate the following equation:

$$Exit_{it} = \zeta_0 + \zeta_1 D_i + \delta_t + xsi_{it}, \tag{3}$$

where $Exit_{it}$ takes 0 if racer i runs at least one race in term t and 1 if the racer runs no race in term t. In this specification, ζ_1 captures the treatment effects on average exit rates over eight years.

the nonlinear regression will cause a bias, while the linear regression will not. Given the power of my regression estimates, I shall choose a linear regression model over nonlinear models that come with great costs in my case. Furthermore, it is not straightforward to estimate interaction terms such as the DID terms in nonlinear models as pointed out by Ai and Norton (2003).

4.4 Characteristics of Dropouts

To shed light on who exits the market, I estimate the outcome differences between the dropouts and remainders at the last terms of the dropouts. Since I do not observe the final-term performance of those dropouts in the end of the eighth year, I will have 15 terms in total. However, since there tend to be generally fewer than 20 dropouts in each term, there would not be enough statistical power to detect the difference. To deal with this issue, I divide the terms into three categories of, term 0 to 4, 5 to 9, and 10 to 14. In particular, I estimate the following equation

$$y^{\tau}_{i\hat{t}} = \kappa^{\tau}_0 + \kappa^{\tau}_1 Last_{i\hat{t}} + \delta_{\hat{t}} + \alpha_i + q_{i\hat{t}}, \tag{4}$$

where $y^{\tau}_{i\hat{t}}$ is racer i's outcome at term $\hat{t} \in \{\tau - 4, 1, ..., \tau\}$ for $\tau \in \{4, 9, 14\}$, $Last_{i\hat{t}}$ is an indicator that equals 1 if $\hat{t}$ is the last term of a dropout *and* i is the dropout, and 0 otherwise. κ^{τ}_1 essentially estimate the mean difference in the outcome variables between the dropouts and remainders within the first five, middle five, and the last five terms.

5 Results and discussions

5.1 Extensive Margin

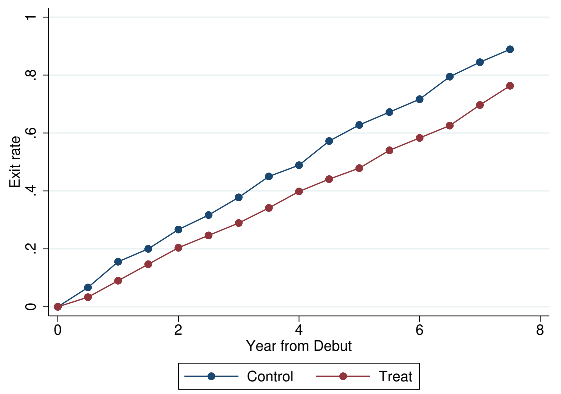
As Figure 2a and 2b show, there appears to be some differences in exit rates in later years between the treatment and control group and also between male and female racers. The average exit rate is lower for the treatment group within the male sample, and the opposite appears to be true within the female sample. This results in some survivorship bias in my estimates, causing a downward bias.

I check the statistical significance of the gap in the (temporary) exit rates from the market in Figure 3 by estimating equation (1). The outcome takes 0 if a racer remains in the market in a specific term and 1 if he or she does not attend any race in that term. As Figure 3a, the exit rate for male racers becomes statistically significantly higher for the control group for some periods as time goes by. The opposite trend is observed among female racers in Figure 3b, although the statistical significance is absent due to the small female sample size within each term.

I also estimate the average effects by estimating equation (3). Table 6 demonstrates that on average, the treatment decreases exit rates by 10% for male racers, while it *increases* exit rates by 6% for female racers. The difference between the male and female racers is also statistically significant. To shed light on the characteristics of these dropouts, I provide the estimation results of equation (4) at Table 7 and 8. For the male sample, we see an interesting pattern in changes in the charactersitics over time. Within the first five terms, we see that the dropouts are those who win fewer races and do not take risks. When racers hit the mid five terms, these racers compete more frequently in lower-graded races

Figure 2: Exit Dynamics

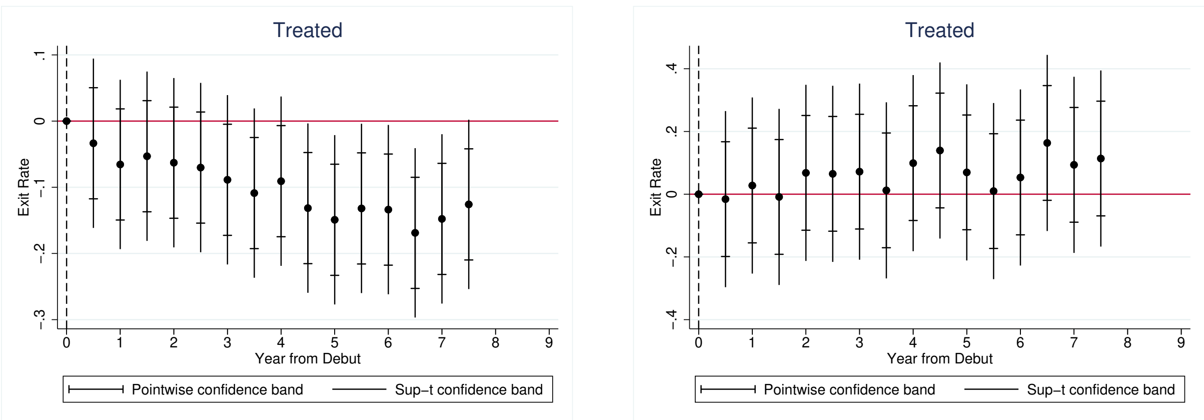
(a) Male (b) Female



Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year.

Figure 3: Exit Dynamics

(a) Male (b) Female



Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year. Coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (1).

without taking risks within these races and earn more than the remainders. This may reflect a strategic behavior of risk-averse racers who secure the third- or fourth-place finishes with no risk rather than aim for the first-place finishes by taking risks. In the last term, this tendency becomes more attenuated except for the risk-taking behavior being inverted perhaps because racers get "admonished" to retire if their performance is poor after their fifth year and get pressured to take risks. On the other hand, we do not see statistical significance in the female sample except for the risk-taking measure, most likely due to the small sample size. The only takeaway in the female sample from the analysis is that those less aggressive female racers tend to exit the market in general.

Table 6: Average Effects on Extensive Margin

	(1) All	(2) Female	(3) Male	(4) Diff
Treat	-0.076***	0.061*	-0.098***	
	(0.010)	(0.024)	(0.011)	
Treat x Male				-0.173***
				(0.027)
N	7424	1168	6256	7424

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. First column uses both male and female racers, the second focuses on female racers, the third column on male racers, and the last column conducts a heterogeneous analysis with fixed effects for the treatment and sex being included.

Table 7: Characteristics of Dropouts: Male Sample

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
First 5 terms				
Last	-0.171*	0.055	-0.032	-0.011**
	(0.067)	(0.042)	(0.035)	(0.004)
N	1712	1712	1712	1712
Mid 5 terms				
Last	0.103	0.118**	-0.230*	-0.027***
	(0.054)	(0.040)	(0.102)	(0.006)
N	1176	1176	1176	1176
Last 5 terms				
Last	0.137**	0.169***	-0.415***	0.015*
	(0.051)	(0.042)	(0.123)	(0.006)
N	659	659	659	659

Notes: Estimation results of equation (4) with the male sample. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8: Characteristics of Dropouts: Female Sample

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
First 5 terms				
Last	-0.080	0.038	-0.012	-0.031**
	(0.155)	(0.095)	(0.049)	(0.011)
N	339	339	339	339
Mid 5 terms				
Last	0.043	0.023	-0.103	-0.034*
	(0.147)	(0.114)	(0.141)	(0.015)
N	211	211	211	211
Last 5 terms				
Last	-0.099	-0.139	-0.288	-0.018
	(0.171)	(0.165)	(0.251)	(0.016)
N	89	89	89	89

Notes: Estimation results of equation (4) with the female sample. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.2 Intensive Margins

5.2.1 Dynamic Effects

In Figure 4 and Figure 5, x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year. Figure 4 plots the mean value of each group for the outcomes of interest. In general, they all show an increasing difference over time. Interestingly, the proportion of inward lane change increases for both groups until year 5.5, but it decreases afterwards. The abrupt drop in the number of GII or above races for the control group in the last term is mainly due to the exit and downgrade of some top racers in the control group. Since the number of racers decreases over time due to the nature of dataset, the later terms are affected more by anomalies induced by a small number of racers. Note that the main results remain unchanged even if I exclude the last term from the analysis.

In Figure 5, coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (1). Panel (a) of Figure 5 corresponds to the total number of the first place, (b) the natural logarithm of total earnings, (c) the total number of GII or above tournament races, and (d) the proportion of the number of races in which racer i changed lanes toward the first lane, all at the term level.

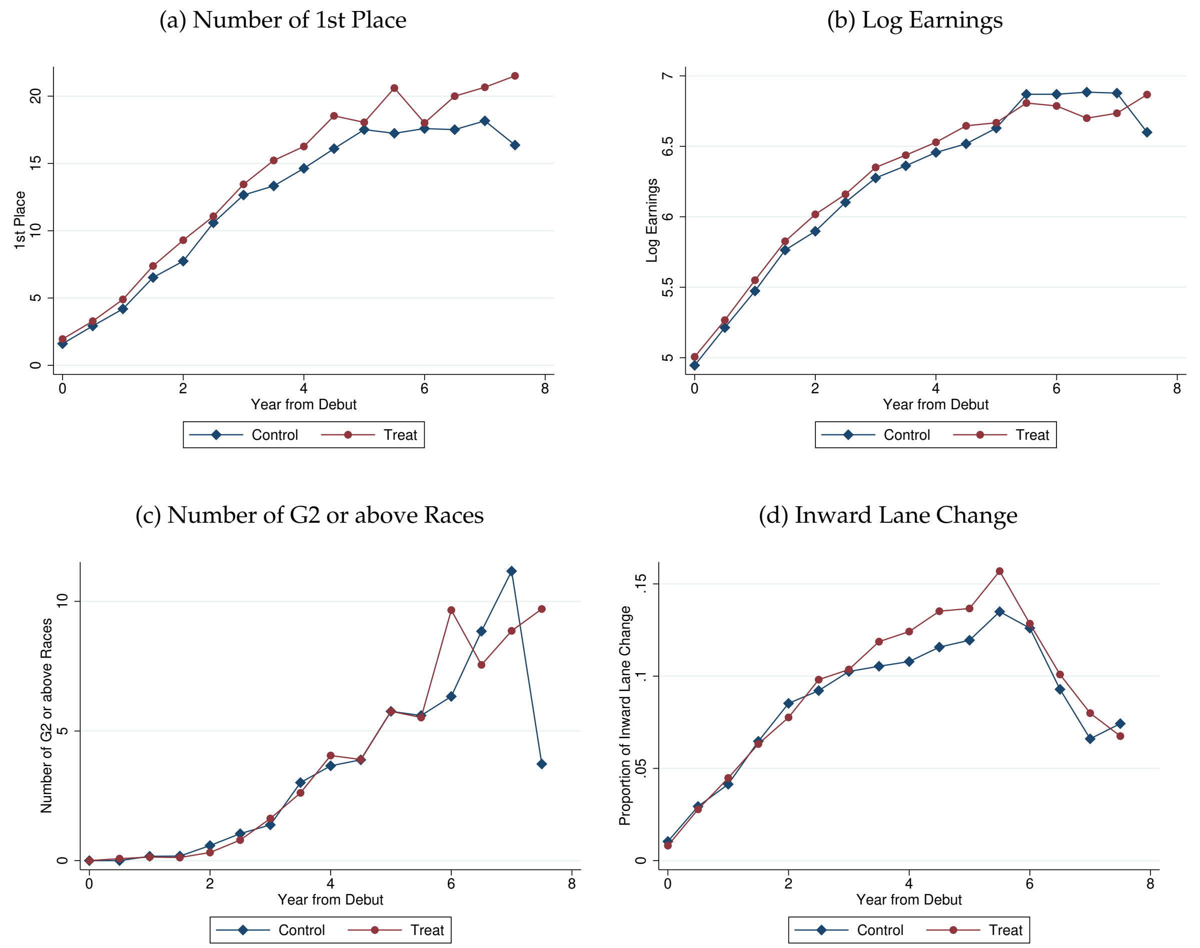
From panel (a), we can see a steady increase in the gap between the treatment and control group in the number of wins, although the statistical significance is barely achieved for some terms pointwise and none of them are if uniform confidence intervals are used. This is because the number of racers are fewer toward the later terms due to the data limitation.

We can see a similar pattern in earnings gap from panel (b). It appears that the percentage gap in earnings widens a bit after 1.5 years. The sudden increase in the mean of the control group after the 5th year comes from exits of low-performing racers in the control group (or the higher retention rates of the treatment group induced by the treatment), which will be analyzed in more depth later. The declines of average earnings in both groups toward the later years come from the decreasing number of racers due to the data limitation. As the sample size decreases toward the end, an exit of a small number of racers affect the average much more, resulting in the amplification of idiosyncratic shocks.

In contrast, we do not see a steady increase in the number of opportunities for the treatment group as portrayed by panel (c). As a racer wins more races, his or her racer grade will be upgraded, which will give the racer qualifications for GII or above-graded tournament races. As mentioned above, only 40% of entire racers can become A2 or A1 grades, and only a handful of racers can attend some GI or SG tournaments. Then, the outcome is highly skewed and hard to detect the effects until later terms. When I analyze the cumulative effects and take the natural logarithm, I overcome this skewness issue with many zeros.

Finally, the dynamics of risk-taking behavior appear to take not as monotonic patterns as the other

Figure 4: Outcome Trend of Male Sample



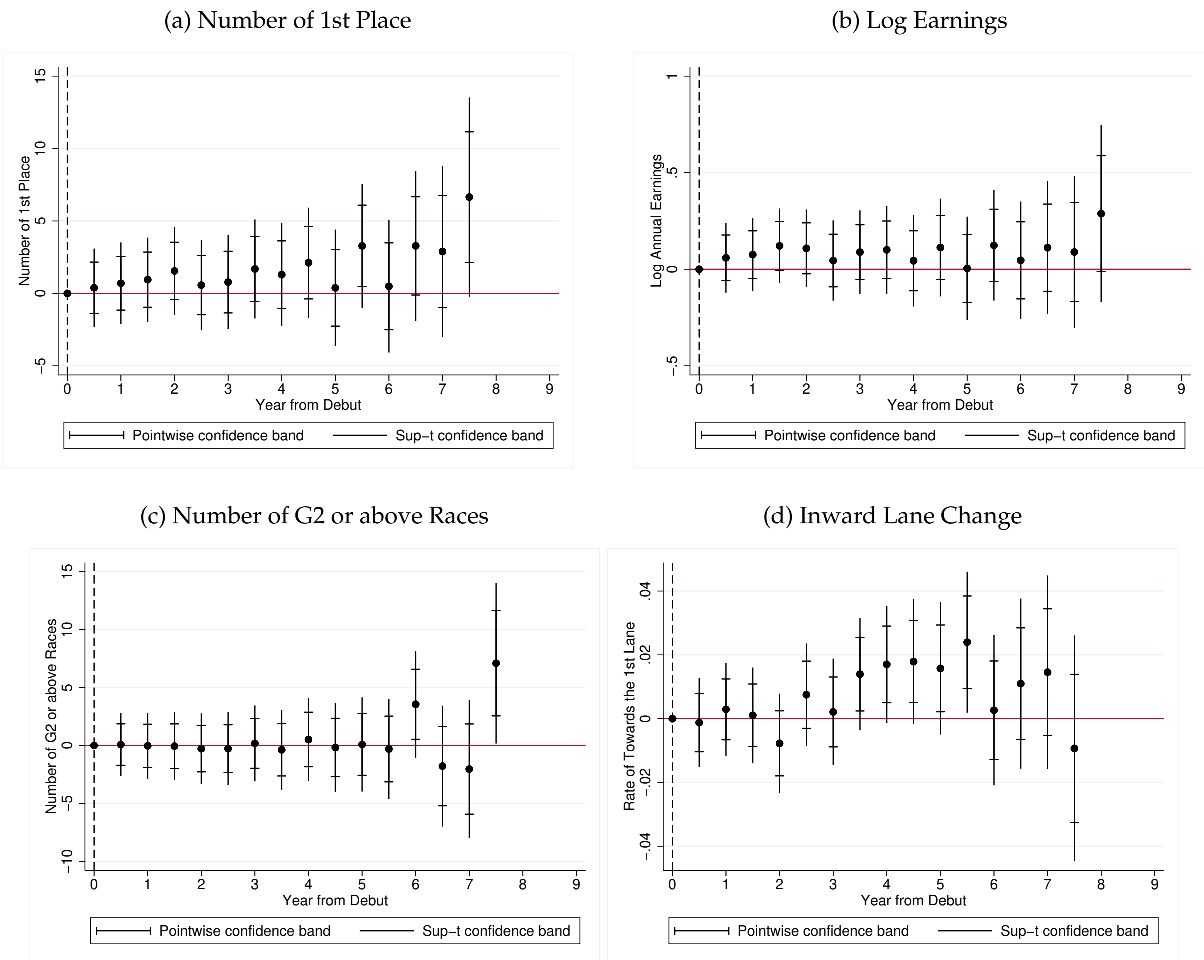
Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year.

outcome as shown in panel (d). Similar to the total number of first places, the proportion of races in which a racer changes the racer's initial lane toward the first lane seems to increase over time, especially until year 6, but the trend inverts afterwards. As demonstrated by the outcome trend, racers start not taking risks once established. The decreasing degree of risk-taking does not seem to come from an institutional reason. The JSRA tends to assign more inward lanes to racers who performed well in preliminaries or higher grade racers. After some years, better racers are assigned the first lane frequently enough to not need to take risks of lane changes toward the first lane during a race. I exclude these observations when I compute the proportion of inward lane change races.

5.2.2 Cumulative Effects

Given the dynamics, I shall estimate the cumulative effects of the initial small differences driven by luck. Table 9 demonstrates these effects. As for the first column, the number of first places increases by approximately 69 percent (100 x [exp(0.522) - 1]). The second column shows approximately a 48 percent

Figure 5: Baseline Results: Male Sample



Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year. Coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (1).

(100 x [exp(0.395) - 1]) difference between the two groups in cumulative earnings. The third column demonstrates around 61% more GII or above tournament races for the treatment group, and the fourth column says 1.2 percent more lane changes toward the first lane, providing evidence for greater risk-taking behavior. These results are robust to the inclusion of controls (listed in the balance table above) that are demonstrated at Table B2 in Appendix B.

Table 9: Cumulative Effects: Male Sample

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.522**	0.395**	0.477**	0.012**
	(0.170)	(0.133)	(0.176)	(0.004)
N	391	391	391	391

Notes: Estimation results of equation (2) with the male sample. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10: Control Group Summary Statistics of Cumulative Outcomes

	mean	sd	min	max
Number of 1st Place	81.09	98.31	0	485.00
Earnings (in 10,000 yen)	3907.63	4853.31	89	34873.89
Number of GII or above Races	17.39	68.94	0	486.00
Proportion of Inward Lane Change	0.06	0.04	0	0.16

5.2.3 Distributional Effects

One may think that the butterfly effects are concentrated the most in the top quantiles of the outcome distributions. To examine this natural conjecture, I conduct simultaneous quantile regressions with 1,000 repetitions of random draws for bootstrapping. Surprisingly, the proportion of the differences driven by the early-stage luck is the largest among the middle-range group. Table 11 demonstrates the distributional effects of the early-stage luck on the cumulative outcomes.

5.2.4 Female Sample

Appendix B shows the results for the female racer sample. Surprisingly, all the positive results are swept away. While the small sample size might be one reason for the absence of the results, the lack of effects appears to come from the fact that female racers are induced to (at least temporarily) exit the market, perhaps for marriage and maternity leave.

As Japan's gender gap is so high that it ranked the 120th among 151 countries in 2021 according to the World Economic Forum, it was not unusual for Japanese women to quit jobs once married during the sample period. The interesting result that those lucky female racers exit the market seems similar to the hump-shaped pattern of wives' working hours by their husbands' wages as documented and

Table 11: Distributional Effects

(a) Number of 1st Place

	(1) q10	(2) q20	(3) q30	(4) q40	(5) q50	(6) q60	(7) q70	(8) q80	(9) q90
Treat	2	4	21**	28*	39***	53***	54*	52	82**
	(1)	(5)	(6)	(12)	(11)	(12)	(21)	(31)	(29)
CMean	1	6	12	26	48	64	94	147	213

(b) Earnings (absolute)

	(1) q10	(2) q20	(3) q30	(4) q40	(5) q50	(6) q60	(7) q70	(8) q80	(9) q90
Treat	82	286	717**	1090*	1515**	1848***	2015**	2445*	2961*
	(153)	(279)	(273)	(538)	(466)	(475)	(727)	(1187)	(1217)
Constant	261	602	1088	1812	2587	3356	4212	6012	8991

Notes: Standard errors in parentheses, whose decimals are rounded up for the table to fit in the paper. Control Mean (CMean) is taken from the values of the constants (intercepts) in simultaneous quantile regressions. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

discussed in e.g., Juhn and Murphy (1997) and Bredemeier and Juessen (2013). In other words, the treatment increased the earnings of female racers which in turn increased the probability of these female racers marrying higher-wage men in the marriage market compared to the control group female racers.

5.2.5 Sample after 2012

As demonstrated in section 3.2, after 2012, an engine becomes a more noisy predictor of a racer performance due to the rule change in propellers, and a new model of an engine since December 2014 makes it even harder for new racers to win a race by a lucky motor. Given this, a natural hypothesis is that the same analysis with 2012 would spit a noisy estimate of the early luck that is weak in magnitude. Table 12 and 13 support this hypothesis. All the point estimates are much smaller than the ones with the main sample. The statistical significance is gone due to the large noise, although the confidence intervals do not cross zero in general.

Table 12: Average Effects on Extensive Margin of Those Who Debuted after 2012

	(1) All	(2) Female	(3) Male	(4) Diff
Treat	-0.013	0.016	-0.018	
	(0.009)	(0.018)	(0.010)	
Treat x Male				-0.035
				(0.020)
N	9408	2320	7088	9408

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample restricted to those racers who debuted after April 30, 2012. First column uses both male and female racers, the second focuses on female racers, the third column on male racers, and the last column conducts a heterogeneous analysis with fixed effects for the treatment and sex being included.

Table 13: Cumulative Effects on Those Who Debuted after 2012

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.188	0.110	0.273	0.000
	(0.147)	(0.093)	(0.177)	(0.001)
N	443	443	443	443

Notes: Estimation results of equation (2) with the male sample restricted to those racers who debuted after April 30, 2012. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.3 Robustness checks

5.3.1 Different cutoffs

As explained above, I chose the 85th percentile as the cutoff for the treatment, so that the sample sizes of the treatment and control groups balance. To check the sensitivity of this cutoff, I conduct robustness checks by different cutoffs of 80th and 90th percentiles. As demonstrated in Table B3 and B4 in Appendix B, the results remain unchanged under the new definitions of the treatment with the two different cutoffs.

5.3.2 Unlucky engines

If lucky engines culminate in the positive feedback loop effects, it is natural to think that unlucky engines would cause negative feedback loop effects. To check if the effects of engines are symmetrically distributed, we conduct the same analyses of section 3.2 with unlucky engines. I define unlucky engines as those engines below some percentiles. Table 14 demonstrates that the effects of unlucky engines are much smaller in magnitude than lucky engines and appear to be non-linear. *lucky_10* is an indicator for an engine whose the first and second-place finish rates are below the 10th percentile, *lucky_20* is that with the 20th percentile, and so forth. The coefficients -0.0058 are less than half of the coefficient of *lucky_90*, 0.0160 in the absolute term. On the other hand, the magnitudes of *lucky_40* and *lucky_60* are similar. This implies that the effects of engines are asymmetric in tails.

This asymmetry appears to come from the model of engines. The new model after December 2014 is more symmetric in tails as demonstrated in Table 15.

Consistent with the results, the cumulative effects of unlucky engines shown in Table 16 are noisy but negative in much smaller magnitude than the lucky engines with the main sample. I use the 15th percentile as the cutoff for the definition of an unlucky treatment and define as the treatment group those new racers who were assigned the unlucky engines within the first five tournaments. Table 16 shows noisy estimates that are negative and statistically insignificant, although the confidence intervals do not cross zero except the measure for opportunities.

Furthermore, Table 17 demonstartes the results of extensive margin analyses of the unlucky treatment. While the magnitudes are smaller and the estimates are noisier, the effects are statistically significant and the signs are opposite of the lucky treatment.

Table 14: Effects of unlucky engines on the probability of the first-place finish by new racers

	(1) firstplace	(2) firstplace	(3) firstplace	(4) firstplace
lucky_10	-0.0058			
	(0.0030)			
lucky_20		-0.0058*		
		(0.0023)		
lucky_30			-0.0061**	
			(0.0020)	
lucky_40				-0.0071***
				(0.0019)
Control Mean	0.0300	0.0307	0.0313	0.0323
N	32817	32817	32817	32817

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Effects of unlucky engines on the probability of the first-place finish by new racers debuted after 2012

	(1) firstplace	(2) firstplace	(3) firstplace	(4) firstplace
lucky_10	-0.0044***			
	(0.0013)			
lucky_20		-0.0034***		
		(0.0010)		
lucky_30			-0.0023**	
			(0.0009)	
lucky_40				-0.0023**
				(0.0008)
Control Mean	0.0071	0.0073	0.0074	0.0076
N	38185	38185	38185	38185

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Cumulative Effects of Unlucky Engines

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	-0.183	-0.154	-0.174	-0.005
	(0.173)	(0.135)	(0.178)	(0.004)
N	391	391	391	391

Notes: Estimation results of equation (2) with the male sample. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 17: Extensive Margin Analysis of Unlucky Treatment

	(1) All	(2) Female	(3) Male	(4) Diff
Treat	0.022*	-0.068**	0.040***	
	(0.010)	(0.023)	(0.011)	
Treat x Male				0.117***
				(0.027)
N	7424	1168	6256	7424

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. First column uses both male and female racers, the second focuses on female racers, the third column on male racers, and the last column conducts a heterogeneous analysis with fixed effects for the treatment and sex being included.

References

AI, C. AND E. C. NORTON (2003): "Interaction terms in logit and probit models," *Economics letters*, 80, 123–129.

ALMOND, D. (2006): "Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population," *Journal of political Economy*, 114, 672–712.

ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

BAILEY, M. J., S. SUN, AND B. TIMPE (2021): "Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency," *American Economic Review*, 111, 3963–4001.

BEDARD, K. AND E. DHUEY (2006): "The persistence of early childhood maturity: International evidence of long-run age effects," *The Quarterly Journal of Economics*, 121, 1437–1472.

BELLÉGO, C., D. BENATIA, AND L. PAPE (2022): "Dealing with logs and zeros in regression models," *arXiv preprint arXiv:2203.11820*.

BOOTH, A. AND E. YAMAMURA (2018): "Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan," *Review of Economics and Statistics*, 100, 581–593.

BREDEMEIER, C. AND F. JUESSEN (2013): "Assortative mating and female labor supply," *Journal of Labor Economics*, 31, 603–631.

CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *oxford Economic papers*, 57, 191–261.

DU, Q., H. GAO, AND M. D. LEVI (2012): "The relative-age effect and career success: Evidence from corporate CEOs," *Economics Letters*, 117, 660–662.

FUJINO, T.-I. (2006): *Yokuwakaru Kyotei no Subete*, Tokyo: Saneki Books.

GENDA, Y., A. KONDO, AND S. OHTA (2010): "Long-term effects of a recession at labor market entry in Japan and the United States," *Journal of Human resources*, 45, 157–196.

GINSBURGH, V. A. AND J. C. VAN OURS (2003): "Expert opinion and compensation: Evidence from a musical competition," *American Economic Review*, 93, 289–296.

HECKMAN, J. J. AND S. MOSSO (2014): "The economics of human development and social mobility," *Annu. Rev. Econ.*, 6, 689–733.

HIMURA, K.-I. (2015): *Yokuwakaru Boat Race no Subete (Basic Knowledge about Boat Race)*, Tokyo: Saneki Books.

——— (2017): *Boat Race no Subetegawakaru (from Introduction to Betting Techniques)*, Tokyo: Saneki Books.

——— (2022): "Motor Seinousa wa Kaishodekinai," [Online article published by a website operated by Makuuru Magazine at https://sp.macour.jp/columns/himura/160519/; posted 6-June-2022].

HOUMARK, M. A., V. RONDA, AND M. ROSHOLM (2024): "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills," *American Economic Review*, 114, 385–425.

JUHN, C. AND K. M. MURPHY (1997): "Wage inequality and family labor supply," *Journal of labor economics*, 15, 72–97.

KAWAGUCHI, D. (2011): "Actual age at school entry, educational outcomes, and earnings," *Journal of the Japanese and International Economies*, 25, 64–80.

MACCINI, S. AND D. YANG (2009): "Under the weather: Health, schooling, and economic consequences of early-life rainfall," *American Economic Review*, 99, 1006–1026.

MULLER, D. AND L. PAGE (2016): "Born leaders: political selection and the relative age effect in the US Congress," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 179, 809–829.

SANDEL, M. J. (2020): *The tyranny of merit: What's become of the common good?*, Penguin UK.

YAMAGUCHI, S., H. ITO, AND M. NAKAMURO (2023): "Month-of-birth effects on skills and skill formation," *Labour Economics*, 102392.

Appendices

A Earnings Data Details

In this section, I note the details of how I computed earnings. Unfortunately, the JSRA does not list the prize amounts and Himura (2015) and Himura (2017) do not list them other than the winner prize amount of the winner of graded tournaments. The main source of prizes for graded tournaments and a lower bound for non-graded tournaments come from two online articles written by two speedboat racing predictors.[18] I fill in the incomplete list described in the data section.

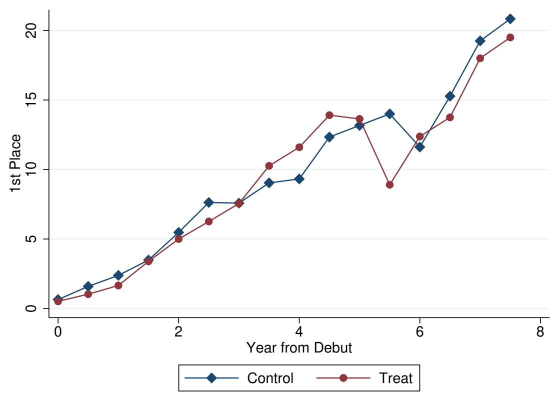
[18] The main source is https://kyoutei-navi.com/sg/grade/ supplemented by https://boat-race.biz/4788 since the first source lacks prize amounts for some within-tournament races such as "dream match" races. Both are accessed on November 13, 2023.

B Additional Results and Robustness Checks

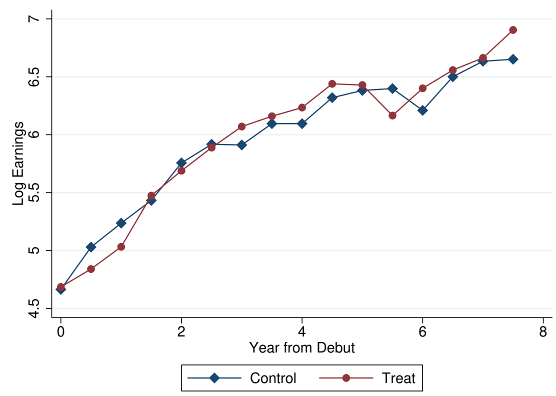
B.1 Female Sample

Figure B1: Outcome Trend of Female Sample

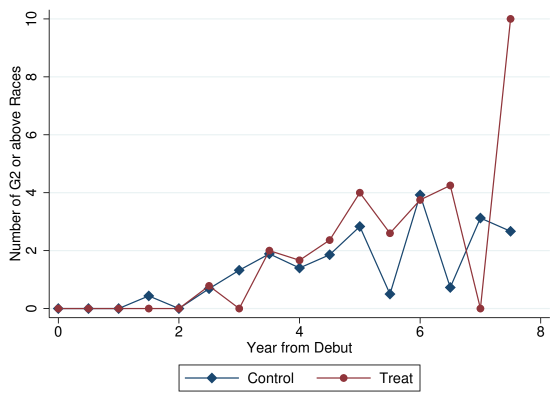
(a) Number of 1st Place



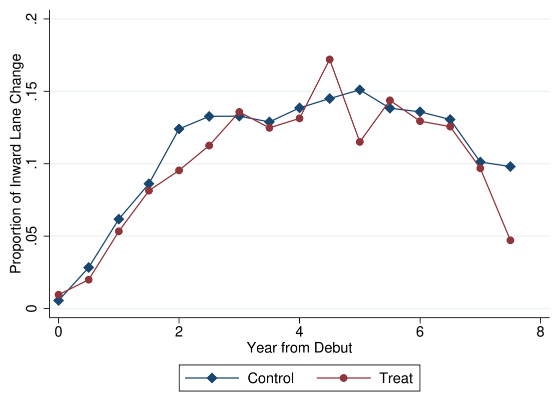
(b) Log Earnings



(c) Number of G2 or above Races

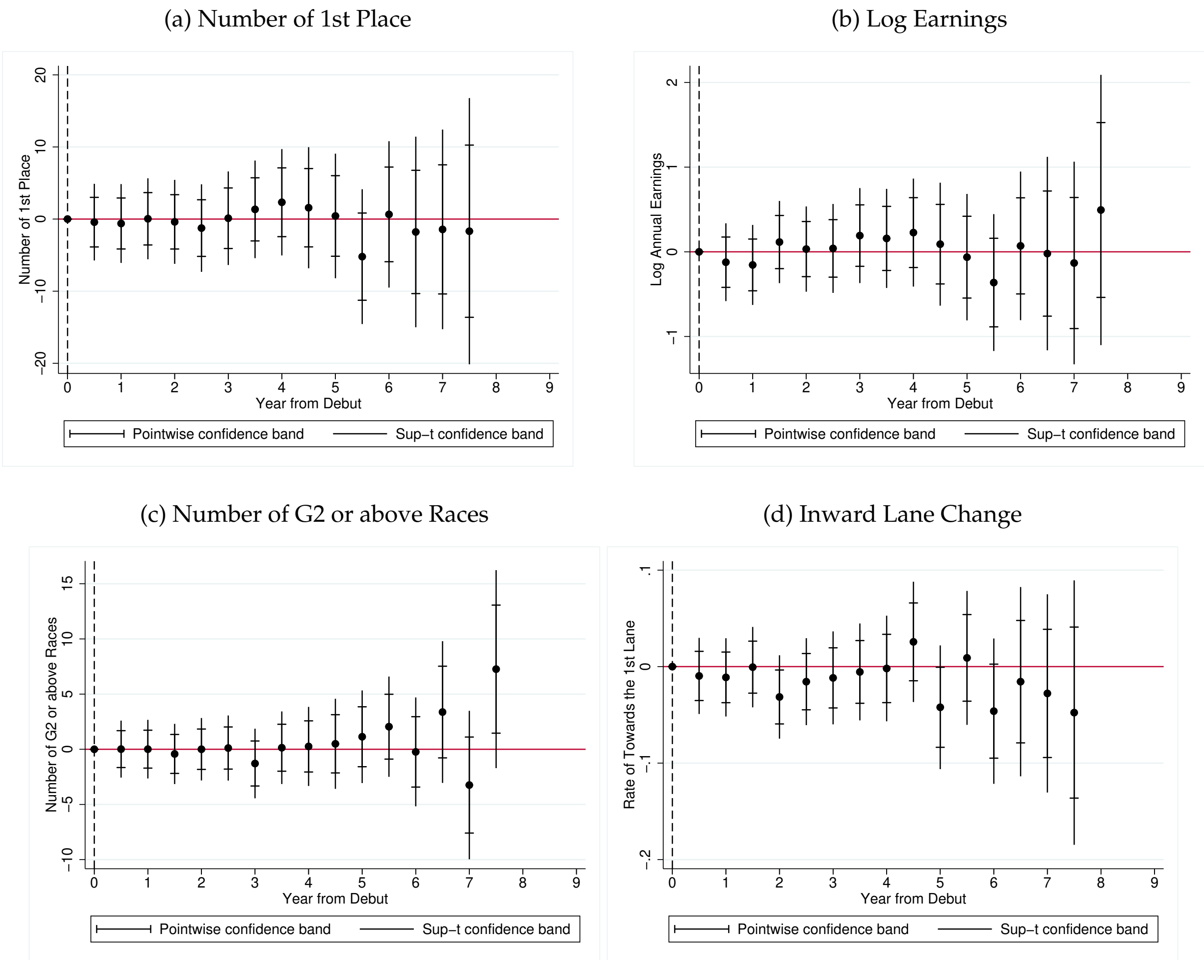


(d) Inward Lane Change



Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year.

Figure B2: Baseline Results: Female Sample



Notes: x-axis indicates time period where 0 is equal to the debut term, and each tick corresponds to a term, so that two ticks equal one year. Coefficients with 95% confidence intervals and uniform sup-t confidence intervals from equation (1).

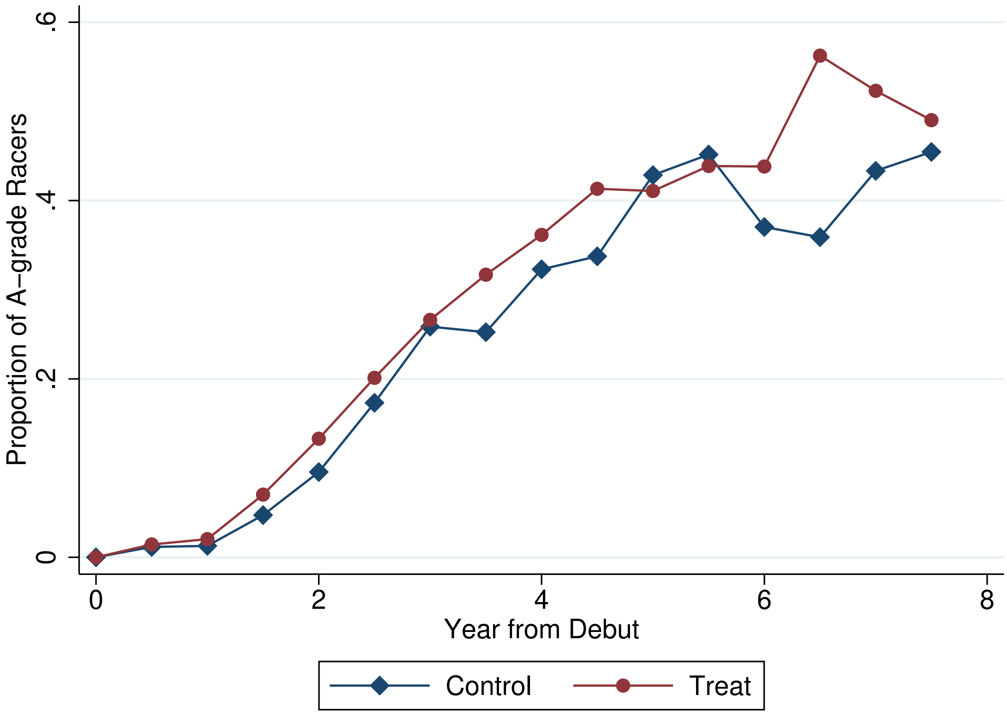
Table B1: Cumulative Effects: Female Sample

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.005	0.043	-0.022	-0.008
	(0.400)	(0.277)	(0.338)	(0.011)
N	76	76	76	76

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B.2 Proportion of Racers whose Racer Grade Is above A2

Figure B3: Proportion of Racers with Racer Grade above A2



B.3 Robustness checks

B.3.1 Main regression with controls

Table B2: Cumulative Effects with Controls: Male Sample

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.399**	0.300**	0.371*	0.010**
	(0.144)	(0.111)	(0.164)	(0.004)
N	391	391	391	391

Notes: Estimation results of equation (2) with controls for the male sample. Controls include (i) height, (ii) weight, (iii) age, (iv) indicator for blood type B, (v) indicator for being born in April or May, and (vi) the winning rate at the Yamato School as the pre-debut performance measure. First column corresponds to the natural logarithm of the number of first-place finishes, the second the natural logarithm of earnings, the third the natural logarithm of the number of races of GII or above grades, and the fourth the proportion of inward lane change conducted. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B.3.2 Different cutoffs

Table B3: Cumulative Effects: 80th percentile

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.550**	0.388**	0.268	0.010*
	(0.178)	(0.140)	(0.186)	(0.004)

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B4: Cumulative Effects: 90th percentile

	(1) 1st Place	(2) Log Earnings	(3) GII or above Races	(4) Inward Lane Change
Treat	0.462**	0.349*	0.456*	0.008*
	(0.176)	(0.138)	(0.182)	(0.004)

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.