



## Data Article

# Dataset on PGA Tour tournament entry

Inés Alegre<sup>a</sup>, Miguel A. Canela<sup>a</sup>, David Pastoriza<sup>b,\*</sup>

<sup>a</sup> IESE Business School, Spain

<sup>b</sup> HEC Montreal, Canada

## ARTICLE INFO

### Article history:

Received 25 November 2021

Revised 19 January 2022

Accepted 9 February 2022

Available online 15 February 2022

### Keywords:

Self-selection

Tournaments

Momentum

Heuristics

PGA Tour

Golf

## ABSTRACT

This article describes a dataset that allows to explore the determinants and moderators of athletes' decision to enter in tournaments endowed with a monetary prize. Specifically, the dataset contains variables that describe athlete's short-term momentum (i.e., performance streak in the tournaments recently entered) and long-term momentum (i.e., performance streak in the same tournament across seasons), which permits an in-depth analysis of how past performance trajectory drives self-selection into tournaments. The dataset consists of 54,915 self-selection decisions that golfers have taken over an eleven-year period (1996–2006) when deciding to participate in PGA Tour tournaments.

© 2022 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

DOI of original article: [10.1016/j.joep.2021.102414](https://doi.org/10.1016/j.joep.2021.102414)

\* Corresponding author.

E-mail address: [david.pastoriza@hec.ca](mailto:david.pastoriza@hec.ca) (D. Pastoriza).

<https://doi.org/10.1016/j.dib.2022.107952>

2352-3409/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Behavioural Finance and Economics
Specific subject area	Behavioural Economics; Sports Economics; Heuristics
Type of data	Table; Figure; Excel file
How data were acquired	PGA Tour data was acquired through licensing agreement with PGA Tour, which allows to use data for scientific purposes. Official World Golf Ranking data was extracted from the institution's website. Players' biographic information was manually extracted from PGA Tour's media guides
Data format	Mixed (raw and processed)
Description of data collection	We accessed to the original ShotLink® data through a contractual agreement with the PGA Tour
Data source location	The data was gathered from the PGA Tour ShotLink® Database and the Official World Golf Ranking.
Data accessibility	It is provided in the supplementary material of this article. Alternatively, to access the data you may enter in <a href="https://www.doi.org/">https://www.doi.org/</a> and introduce the following code: <a href="https://doi.org/10.17632/ywtwg39x3x.1">10.17632/ywtwg39x3x.1</a>
Related research article	Pastoriza, D., Alegre, I., & Canela, M. A. (2021). Conditioning the effect of prize on tournament self-selection. <i>Journal of Economic Psychology</i> , 86, 102,414.

Value of the Data

- The increasing ubiquity of prize-based contests has been widely acknowledged by industry analysts. This situation urges the public and private sectors to understand how tournament design can contribute to attract the best possible contenders. The dataset in this article contributes to our understanding of what drives self-selection into tournaments.
- Thus far, research on tournament self-selection has not accounted for the performance momentum that agents have had prior to entering a tournament [1]. This dataset is useful for researchers who want to understand how past momentum [2] influences tournament entry.
- The dataset distinguishes between within-season and across-season momentum. This allows researchers to examine the effect of short-term and long-term momentum on tournament entry.
- The dataset can be used to further understand heuristics [3,4] and how the latter helps agents to interpret winning and losing streaks; that is, how positive and negative momentum influence agents' expectations about future performance.
- The dataset can be used to understand the downsides of heuristics, such as leading individuals to act on erroneous biases when, for instance, they erroneously self-selecting into tournaments based on their positive transitory momentum (i.e., even when that transitory momentum does not result into positive subsequent performance).

1. Data Description

The ubiquity of prize-based contests has been widely acknowledged by industry analysts [5]. Despite the progress made thus far, research on tournament self-selection has not accounted for the performance momentum that agents have had prior to entering a tournament. However, in reality an agent's performance is dynamic – i.e., two agents may arrive at a tournament from differing performance trajectories (i.e., one is on a positive streak while the other is one a negative streak). Little is known about how past performance trajectory influences tournament self-selection [6], partly due to a lack of data availability.

The dataset provided in this article precisely allows for a comprehensive analysis of the influence of momentum on tournament entry. The database not only disaggregates momentum into both short-term and long-term, but also into positive and negative streak. Additionally, it provides a series of variables that could drive/moderate the tournament self-selection decision, and that the researcher may want to account for. The following table describes the list of variables in the dataset:

**Table 1**

Description of the variables in the dataset.

Variable	Type	Description
PlayNum <sup>1</sup>	Numeric	Player Identification number
Season	Numeric	Yearly season
TournamentChronologicalOrder	Ordinal	Chronological order in which the tournaments took place within the season
TournamentEntered	Categorical {1; 0}	Binary variable that indicates whether the player entered the tournament. Takes value 1 if the player entered the tournament and 0 otherwise
TournamentPrizeMoney	Numeric	Prize of the tournament in US dollars
CompetitivenessTournament	Numeric	Measure of the aggregated level of ability of the players who entered the tournament. It is calculated using the formula available on the site of the Official World Golf Ranking
InvitationalAlternateTournament	Categorical {1; 0}	Takes value equal to 1 when the regular tournament is an alternate or invitational event, which is less prestigious than a regular tournament (0)
AbilityRanking	Numeric	Ability of the player. It is determined by his position in the Official World Golf Ranking
CumulativeCareerMoney	Numeric	On-the-course money that the player has accumulated on the PGA Tour tournaments (going back as early as 1983)
LongInjurySeason	Categorical {1; 0}	Dummy indicating whether the player is injured by the time he enters a tournament
Temperature	Numeric	Average temperature during the four rounds of the tournament. Meteorological conditions are known for influencing tournament self-selection
DistanceHomeTournament	Numeric	Distance (kilometers) between the residence of the player (updated by season) and the tournament.
DifferenceDistanceHomeTournament	Numeric	Geographical distance (kilometres) between the current tournament and the next one
Distance to Losing PGA Card <sup>2</sup>		
ExemptNextSeason	Categorical {1; 0}	Takes value equal to 1 if the player won a tournament in the current season, which grants him exempt status for next season (i.e., right to enter PGA Tour tournaments next season)
PercentageMoneyLeft	Numeric	Percentage of the current season's total prize money that remains to be allocated. The lower the percentage remaining, the higher the competitive pressure
DistanceLosingCard	Numeric	Ranks distance (in the season's cumulative money ranking) between his current position and player #125 (i.e., last rank to keep the card). A value of -9 (+9) means that he is 9 ranks below (above) the threshold of survival, and thus in a provisional dangerous (safe) position to keep the card
Players' Performance Momentum <sup>3</sup>		
PositiveMomentum_MadeCut	Numeric	Number of consecutive made cuts <sup>4</sup> that the player has been able to string together in the tournaments recently entered within the season

(continued on next page)

Table 1 (continued)

Variable	Type	Description
NegativeMomentum_madecut	Numeric	Number of consecutive missed cuts that the player strung in the tournaments recently entered within the season
PositiveAcrossSeasonStreak_cut	Numeric	Number of consecutive made cuts that the player strung in the same tournament (e.g., U.S. Open) in the seasons prior to the present season. A tournament non-entry does not break the positive across-period momentum, but an entry with negative performance (i.e., not making the cut) does
NegativeAcrossSeasonStreak_cut	Numeric	Number of consecutive missed cuts that the player strung in the same tournament (e.g., U.S. Open) in the seasons prior to the present season. A tournament non-entry does not break the negative across-period trajectory. When the player makes the cut in that tournament, the negative across-period trajectory resets to zero

<sup>1</sup> With the PlayNum, the researcher can identify the player's name on PGA Tour's website.  
<sup>2</sup> This ensemble of three variables allow to capture a player's distance to losing his PGA Card license (i.e., right to play in the PGA Tour the next season) and how that may affect his current tournament self-selection strategy.  
<sup>3</sup> This ensemble of four variables allow to capture a player's short-term and long-term performance trajectory.  
<sup>4</sup> PGA Tour tournaments consist of four rounds. Of all players entering the tournament, only the half with the lowest cumulative strokes after two rounds "make the cut," advancing to the final two rounds. Players who do not "make the cut" receive no prize.

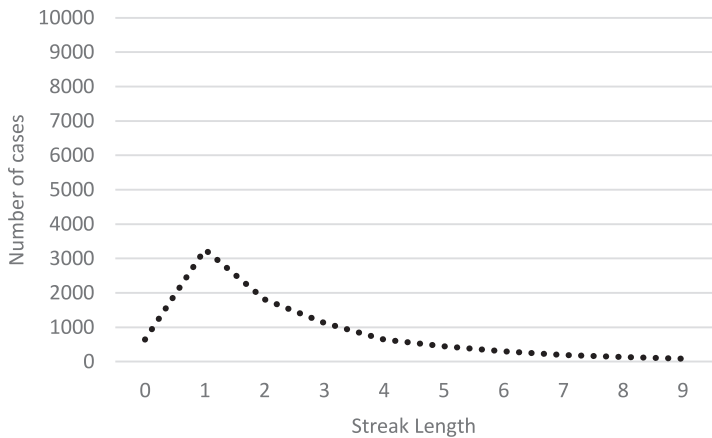
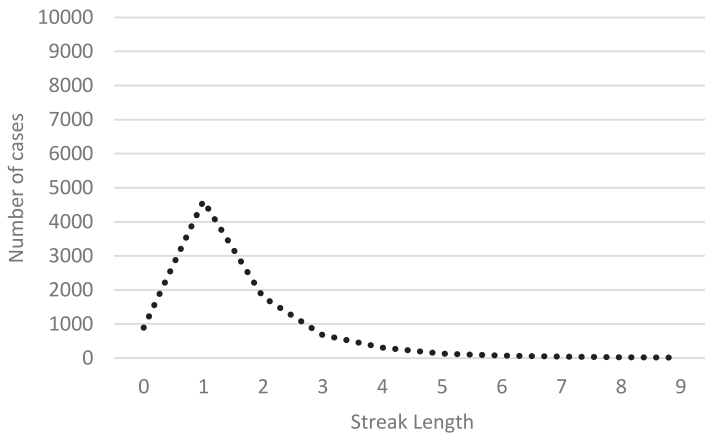
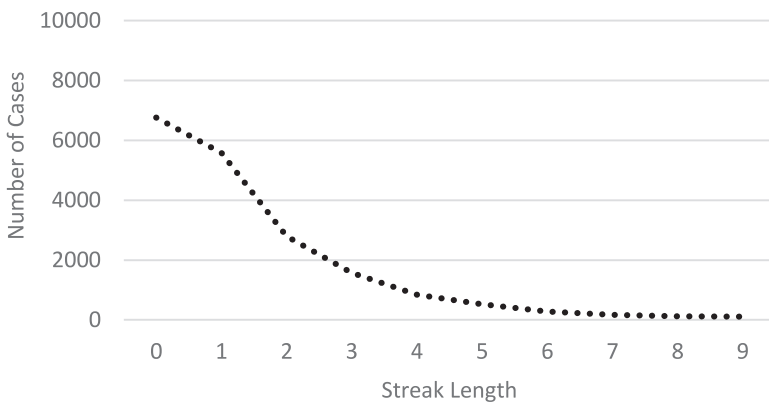


Fig. 1. Frequency of positive within-season momentum.

Figs. 1–3, and 4 below provide describe the momentum variables; specifically, what is the number of observations in the dataset for each streak length. For instance, Fig. 1, which reflects the frequency of positive within-season momentum, shows that there are less than 1000 cases in the sample in which the player had positive within-season streak length equal to 0, a number that increases to over 3000 cases in which the player had a positive within-season streak length equal to 1 (i.e., by the time the player entered a tournament, he had accumulated one consecutive made cut in the previous tournament entered). Similarly, Fig. 4 shows that there are approximately 2000 cases in which the player had a negative across-season streak length equal to 2 (i.e., two consecutive non-made cuts in the same tournament in the same tournament in the seasons prior to the present season).



**Fig. 2.** Frequency of negative within-season momentum.



**Fig. 3.** Frequency of positive across-season momentum.

The dataset is restricted to players who have full-exempt status<sup>1</sup> during a given season. The dataset excludes non-fully exempt players because the latter often merely fill the available spots in a tournament, and thus they cannot plan in which tournaments they will compete. In other words, only exempt players can be strategic about the tournaments they enter. By focusing on individuals who can self-select into all PGA tournaments, we avoid the problem of confounding the sorting effect with the tournament incentive effect [7].

The dataset is restricted to regular tournaments. Non-regular tournaments (i.e., four Major Championships and three World Golf Championships) are excluded because, even though they are part of the PGA Tour schedule, they are the most prestigious and financially rewarding tournaments – i.e., any player who is eligible to enter those tournaments will enter, and as a result there is no self-selection decision. Additionally, non-regular tournaments are excluded because

<sup>1</sup> Approximately 150 players enter each PGA Tour tournament. However, every season there are 250 players per season who own a PGA Tour card, which gives them the right to enter its tournaments. Since there are fewer spots available (150 players) than PGA Tour cards (250 players), each season the PGA Tour creates a ranking that determines who has priority to enter the tournaments. At the top of that ranking are the fully exempt players, who can enter any tournament of their choice.

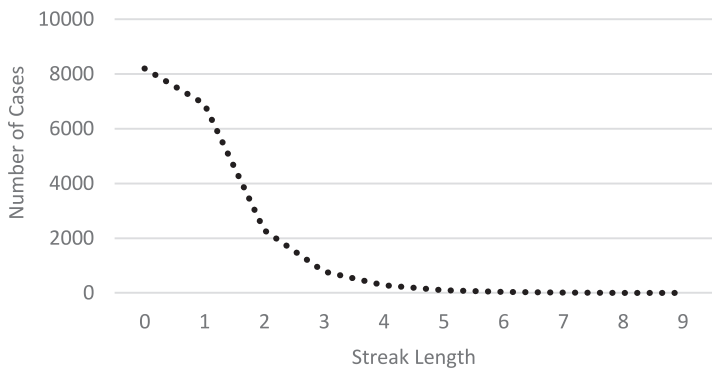


Fig. 4. Frequency of negative across-season momentum.

Table 2  
Descriptive statistics of the variables.

Variable	Mean	S.D.	Min	Max
TournamentEntered	0.57	0.49	0	1
TournamentPrizeMoney <sup>‡</sup>	3.44	1.47	1	8.19
CompetitivenessTournament	317.23	162.77	20	810
InvitationalAlternateTournament	0.24	0.43	0	1
AbilityRanking	89.38	42.72	1	167
CumulativeCareerMoney <sup>‡</sup>	6.15	5.83	0	69.87
LongInjurySeason	0.02	0.15	0	1
Temperature	67.99	9.09	44.20	91
DistanceHomeTournament <sup>‡</sup>	1.91	1.33	0.01	8.18
DifferenceDistanceHomeTournament <sup>‡</sup>	−0.01	1.47	−4.77	7.52
ExemptNextSeason	0.33	0.47	0	1
PercentageMoneyLeft	0.51	0.28	0	0.97
DistanceLosingCard	−31.77	67.29	−124	253
PositiveMomentum_MadeCut	1.86	2.72	0	26
NegativeMomentum_madecut	0.69	1.37	0	23
PositiveAcrossSeasonStreak_cut	1.34	1.93	0	17
NegativeAcrossSeasonStreak_cut	0.53	0.91	0	13

<sup>‡</sup> Expressed in millions of dollars.  
<sup>‡</sup> Expressed in thousands of kilometers.

each of them has a unique entry criteria, and therefore not even the fully exempt players are necessarily exempt.

Our observations start in 1996, since prior to that season it is not possible to identify which players were fully exempt. Our observations end in the 2006 season because in the 2007 season the PGA Tour introduced the FedEx Cup, which is a season-long points contest whereby players who accumulate enough points throughout the season can qualify for a playoff contest in which they compete for more than \$30 million. Naturally, with such a strong monetary incentive, the FedEx Cup could modify players' tournament entry decisions, adding noise to our data.

On Table 2 below you can see the descriptive statistics of the variables in the dataset.

2. Experimental Design, Materials and Methods

Our dataset was computed from two databases. First, we retrieved and computed several tournament-level variables (e.g., tournament entry) from the ShotLink® database, which provides detailed tournament-level information on players' scoring in every tournament, the full access to the ShotLink® database was granted by the PGA Tour via a contractual agreement. This

database allows to trace the cumulative tournament performance of every player at every PGA Tour tournament since 1983. Based on the database information, we coded and created the four performance momentum variables (i.e., positive within-season, negative within-season, positive across-seasons, and negative across-seasons), along with the chronological order of the tournament in the season, whether the player entered the tournament, the tournament prize money, player's cumulative career money, and the player's distance to losing the PGA Tour card. The second database was that of the Official World Golf Ranking database, which provides a measure of players' ability. This second database was obtained directly from the Official World Golf Ranking, which allowed us to calculate the competitiveness of each tournament (i.e., strength of the field of players entering the tournament).

The other variables were retrieved from three different sources. First, the player's health condition (i.e., injuries) and his residence were manually retrieved from PGA TOUR's Media Guides, booklets that the PGA Tour produces for the media containing players' biographic information updated every season. Second, tournaments meteorological conditions were manually retrieved from the site Weather Underground. Third, after identifying the coordinates (i.e., altitude and latitude) of each tournament and the coordinates of each player's residence (updated every season), we computed the distance between locations with the Haversine formula, which is commonly used to calculate the distance between points on the surface of a sphere.

A word of caution is due for those researchers who are unfamiliar with the PGA Tour. Since golfers are independent contractors, they can reject to enter tournaments amongst other reasons because the prize money is not high enough, the distance between their home and the tournament venues is long, or they consider the tournament not to be prestigious enough. As a result, players do not enter in all the tournaments of the season; rather, they plan the tournaments they will enter in the season, often based on their assessment of how their golfing skills fit the attributes of the golf course or when the tournament is placed in the calendar of the season.

## Ethics Statement

The present study did not involve any experiment using human subjects or animals. Thanks to the contractual agreement with the PGA Tour, the latter allows the authors to publish data computed from the ShotLink Database if the publication is in an academic, peer-reviewed journal.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## CRedit Author Statement

**Inés Alegre:** Conceptualization, Data curation, Writing – original draft; **Miguel A. Canela:** Methodology, Software; **David Pastoriza:** Conceptualization, Data curation, Writing – original draft.

## Acknowledgments

We thank the PGA Tour for providing us access to ShotLink® data. We are grateful to David Emond (Delta Statistique) for his coding.

## References

- [1] D. Pastoriza, I. Alegre, M. Canela, Conditioning the effect of prize on tournament self-selection, *J. Econ. Psychol.* 86 (2021) 1–19.
- [2] D.W. Lehman, J. Hahn, R. Ramanujam, B.J. Alge, The dynamics of performance-risk relationship within a performance period: the moderating role of deadline proximity, *Organ. Sci.* 22 (2015) 1613–1630.
- [3] D. Kahneman, A. Tversky, Subjective probability: a judgment of representativeness, *Cogn. Psychol.* 3 (1972) 430–454.
- [4] A. Tversky, A. Kahneman, Judgment under uncertainty: heuristics and biases, *Science* 185 (1974) 1124–1131.
- [5] L. Linnemer, M. Visser, Self-selection in tournaments, *J Econ. Behav. Organ.* 126 (2016) 213–234.
- [6] T. Dohmen, A. Falk, Performance pay and multidimensional sorting: productivity, preferences, and gender, *Am. Econ. Rev.* 101 (2011) 556–590.
- [7] S. Parsons, N. Rohde, The hot hand fallacy re-examined: new evidence from the English Premier League, *Appl. Econ.* 47 (2015) 346–357.