Dynamic Tournament Design: Evidence from Prediction Contests*

Jorge Lemus† Guillermo Marshall‡

May 18, 2020

Abstract

Online contests have become a prominent form of innovation procurement. Contest platforms often display a real-time public leaderboard to provide performance feedback. The impact of information disclosure on players' decisions is theoretically ambiguous: some players may get discouraged and quit, while others may decide to keep working to remain competitive. We empirically investigate the impact of a public leaderboard on contest outcomes using two complementary approaches. First, we compare the equilibria with and without a leaderboard using a dynamic model that we estimate using observational data. Second, we present experimental evidence from student competitions. We find that a real-time public leaderboard improves competition outcomes on average.

Keywords: Dynamic contest, information design, innovation, platforms.
JEL codes: C51, C57, C72, O31.

*We thank Ali Hortaçsu (the Editor) and three anonymous referees for helpful suggestions. We also thank participants and discussants at the Conference on Internet Commerce and Innovation (Northwestern University), IIOC 2017, 3rd Workshop on Structural IO (Cali, 2019), North American Summer Meeting of the Econometric Society (Seattle, 2019), Rob Porter Conference (Northwestern University), Second Triangle Microeconomics Conference (UNC), University of Georgia, Cornell University (Dyson), University of Technology Sydney, and University of Maryland for helpful suggestions. Approval from the University of Illinois Human Subjects Committee, IRB18644.

†University of Illinois at Urbana-Champaign, Department of Economics; jalemus@illinois.edu
‡University of British Columbia, Sauder School of Business; guillermo.marshall@sauder.ubc.ca

1 Introduction

Online competitions have become a valuable resource for government agencies and firms to procure innovations. For instance, U.S. government agencies have sponsored over 1,000 competitions that have awarded over $250 million in prizes for software, ideas, or designs through the website www.challenge.gov.[1] In the UK, the website www.datasciencechallenge.org was created to "drive innovation that will help to keep the UK safe and prosperous in the future." Firms increasingly sponsor competitions on online platforms.[2] One important feature in the design of these platforms is the choice of information disclosure. Some platforms do not disclose information to participants, while others display a real-time public leaderboard with information about the performance of all participants.

Theoretically, displaying a real-time public leaderboard has ambiguous effects on contest outcomes in a dynamic setting, which motivates our empirical analysis. Observing the competition unfolding in real time may discourage some players, but it may also inform other players that more work is required to remain competitive. In contrast, without a public leaderboard, a player must decide whether to continue working or to exit the competition without observing the performance of her rivals. Players do not get discouraged by the realized performance of their rivals without a leaderboard, but the lack of information about their rivals' performance creates uncertainty in their calculation of how much work is required to remain competitive.

Our contribution is to investigate the impact of information disclosure, in the form of a real-time public leaderboard, on contest outcomes. Our setting is Kaggle,[3] an online platform that hosts *prediction contests*, i.e., competitions where the winner is the player with the most accurate prediction of some random variable.[4] Kaggle competitions feature a real-time public leaderboard reporting scores, computed using an objective scoring rule, for all submissions during the competition.

[1]E.g., DARPA sponsored a $500,000 competition to accurately predict cases of chikungunya virus: http://www.darpa.mil/news-events/2015-05-27 (Visited on April, 2019).

[2]Examples include CrowdAnalytix, Tunedit, InnoCentive, Topcoder, HackerRank, and Kaggle.

[3]https://www.kaggle.com/

[4]For instance, IEEE sponsored a $60,000 contest to diagnose schizophrenia; The National Data Science Bowl sponsored a $175,000 contest to identify plankton species from multiple images.

In our analysis, we use two complementary approaches. First, we estimate a dynamic model using observational data from Kaggle competitions, and we use the model estimates to compare the equilibria with and without a public leaderboard. Second, we run a randomized control trial on Kaggle to provide an answer that is independent of our modeling assumptions.

Our continuous-time model captures relevant features of Kaggle competitions. Players can submit multiple solutions but they work on at most one submission at a time. The score of each submission is a random variable drawn from some distribution. A player's type determines the distribution from which scores are drawn. After entering the contest, a player draws a cost from a distribution, which represents the cost of making a new submission. The player then decides to make a new submission or to exit the competition. To make this decision, the player compares the expected payoff of a new submission minus its cost versus the payoff of reaching the end of the competition with her current set of submissions. The player internalizes that her scores may influence other players' decisions through the information released in the public leaderboard. That is, the player understands that achieving a high score may discourage players in what remains of the competition. The player is forward looking in that she takes into account that she and other players may play again in the future. If the player makes a new submission, then she works on that submission (and only that submission) for a random amount of time. When the submission is completed it is immediately evaluated and its score is posted on the public leaderboard.[5] At this point, and after observing the current state of the public leaderboard, the player draws a new cost and again decides to make a new submission or to quit.

We use data from 57 large Kaggle competitions to estimate the primitives of our model, which we do using an estimator based on conditional choice probabilities (CCPs) (Hotz and Miller, 1993; Hotz et al., 1994). While a full-solution method where the econometrician finds the equilibrium of the dynamic game with a real-time leaderboard for every trial vector of parameters is in principle possible, the high dimensionality of the state space makes this approach computationally infeasible. In our approach, we first

[5]The distributions of random variables for submission costs, scores, and submission completion times are common knowledge. We do not model the choice of keeping a submission secret. As we explain in Section 2, the evidence does not indicate strategic behavior in the timing of submissions.

estimate the CCPs using data on decisions as well as data on the state variables faced by the players at the moment of making those decisions. For every observation in our estimation sample, we then simulate the expected payoffs associated to each action in the choice set (i.e., make a new submission or quit the competition) given the state variables faced by the player. In this step, we use the estimates of the CCPs to capture the player's beliefs about how all players would behave in what remains of the competition in response to the player's action. Lastly, we write a likelihood function based on these expected payoffs, and estimate the parameters using a maximum likelihood estimator.

We recompute the equilibrium of the game in the counterfactual scenario where the leaderboard is not displayed. The equilibrium of this game is the solution of an optimal stopping time for each player (similar to Taylor, 1995). We study the impact of the leaderboard on contest outcomes by comparing several outcomes across the equilibria with and without a leaderboard: the total number of submissions, the number of submissions by player type, and the maximum score. We find that a public leaderboard has an economically significant and positive effect on both the number and the quality of the submissions. With a public leaderboard, the number of submissions increases by 21 percent on average, which is explained mostly by an increase in the number of submissions by high-type players. Consistent with this finding, the maximum score on average increases by 1.7 percent with a leaderboard.

The impact of displaying a real-time leaderboard on contest outcomes is heterogeneous across different competitions, depending on the primitives of each competition. In line with theoretical results, we find that the cost of making a new submission relative to the prize (the cost-to-prize ratio) and the variance of the distribution of scores of high-type players are positively correlated with the difference in the number of submissions with and without a leaderboard. First, when the cost-to-prize ratio is high, the cost of erring on the side of staying in the contest for too long is high, so players with no information stop playing too early. Second, a large variance in the distribution of scores of high-type players changes the competition leader more frequently, and we show that players are on average encouraged by a real-time public leaderboard in this case. This suggests that a contest designer would benefit from displaying a public leaderboard in competitions where there is a significant variation in players' scores—e.g., when there are many different approaches to solving a problem—or where new submissions are costly relative to the prize.

Finally, we use a randomized control trial to provide a complementary answer to the question of how performance feedback impacts contest outcomes. To this end, we created and hosted 44 student competitions on Kaggle.[6] Half of the competitions were randomly assigned to the control group (i.e., no public leaderboard) and the other half were assigned to the treatment group (i.e., public leaderboard), with competitions being otherwise equal. The experimental results show that displaying a public leaderboard has a significant and positive effect on both the number of submissions and the maximum score. These "model free" results provide further evidence that a public leaderboard improves competition outcomes on average.

1.1 Related Literature

Contests are a widely used open-innovation mechanism (Chesbrough et al., 2006). They attract talented individuals with different backgrounds (Jeppesen and Lakhani, 2010; Lakhani et al., 2013) and procure a diverse set of solutions (Terwiesch and Xu, 2008).

An extensive literature on static contests has focused on design features such as the number and allocation of prizes, and the number of participants. Studies on the optimal allocation of prizes include the work of Lazear and Rosen (1979), Taylor (1995), Moldovanu and Sela (2001), Che and Gale (2003), Cohen et al. (2008), Sisak (2009), Olszewski and Siegel (2015), Kireyev (2016), Xiao (2016), Strack (2016), and Balafoutas et al. (2017). This literature, surveyed by Sisak (2009), has found that the convexity of the cost of effort plays an important role in determining the optimal allocation of prizes. Taylor (1995) and Fullerton and McAfee (1999), among others, show that restricting the number of competitors in winner-takes-all tournaments increases the equilibrium level of effort. Intuitively, players have less incentives to exert costly effort when they face many competitors, because they have a smaller chance of winning.

The role of information disclosure in dynamic settings has only recently been explored. Rieck (2010) studies a contest where players pay a cost to draw scores from a distribution

[6] All of the participants were students at the University of Illinois at Urbana-Champaign.

and compares participation and the maximum score with and without a leaderboard. The main finding is that the effect of a leaderboard on contest outcomes is ambiguous, and it depends on the primitives of the contest such as costs and the distribution of scores. In a different setting, Aoyagi (2010) compares the provision of effort by agents in a dynamic tournament under full disclosure of information (i.e., players observe their relative position) versus no information disclosure. Ederer (2010) adds private information to this setting whereas Klein and Schmutzler (2016) add different forms of performance evaluation. Goltsman and Mukherjee (2011) study when to disclose workers' performance. Other recent theoretical articles studying dynamic contest design include Halac et al. (2014), Bimpikis et al. (2014), Benkert and Letina (2016), and Hinnosaar (2017).[7]

A growing empirical literature on contests includes Boudreau et al. (2011), Genakos and Pagliero (2012), Takahashi (2015), Boudreau et al. (2016), Bhattacharya (2016) and Zivin and Lyons (2018). Gross (2015) studies how the number of participants changes the incentives for creating novel solutions versus marginally better ones. In a static environment, Kireyev (2016) uses an empirical model to study how elements of contest design affect participation decisions and the quality of outcomes. In his model, players decide up-front how many submissions to send to the contest, i.e., decisions are not based on dynamic information revelation as in our setting. Huang et al. (2014) estimates a dynamic structural model to study individual behavior and outcomes in a platform where individuals can contribute ideas, some of which will be implemented at the end of the contest. Their paper focuses on learning the value of ideas rather than on contest design. Finally, Gross (2017) studies how performance feedback impacts participation in design contests, but the analysis abstracts away from the dynamics of competition. Stopping decisions are based on each players' past outcomes and not on a dynamic leaderboard. This is in contrast with our paper, where we allow for sequential participation and dynamic feedback based on other competitors' performance.

The "gamification" literature—which studies the application of game-design elements (e.g., leaderboards) to areas such as education, marketing, health, or labor markets,

[7]Design levers other than prizes, limited entry, or feedback have been studied, for instance by Megidish and Sela (2013) (requiring a minimal level of effort to participate) and by Moldovanu and Sela (2006) (splitting competitors into two divisions).

among others—is also related. Most of these articles conduct experiments. Landers and Landers (2014) show that adding a leaderboard improves "time-on-task" in a education setting. Landers et al. (2017) show that a leaderboard motivates agents to set more ambitious goals. Athanasopoulos and Hyndman (2011) find that a leaderboard improves forecasting accuracy.

The literature on effort provision for non-pecuniary motives is also related. Lerner and Tirole (2002) argue that high-quality contributions are a signal of ability to potential employers. Moldovanu et al. (2007) studies a setting where status motivates participation. Finally, it is possible to establish a parallel between a contest and an auction. While there is a well-established empirical literature on bidding behavior in auctions (Hendricks and Porter, 1988; Li et al., 2002; Bajari and Hortacsu, 2003, among others), there are only a few papers analyzing dynamic behavior in auctions (see, e.g., Barkley et al., 2019; Coey et al., 2019), which are closer to our work.

2 Background, Data, and Motivating Facts

2.1 Background and Data

We use publicly available information on 57 featured competitions hosted by Kaggle.[8] These competitions offered a monetary prize that ranged between $1,000 and $500,000 (and averaged $30,489), received at least 1,000 submissions from an average of 894 teams per contest, and evaluated submissions according to a well-defined rule. A partial list of competition characteristics are summarized in Table 1 (see Table A.1 in the Online Appendix for the full list).

Participants of Kaggle competitions have access to a training and a test dataset. An observation in the training dataset includes both an outcome variable and covariates; while the test dataset only includes covariates. A valid submission in a contest must include an outcome variable prediction for each observation in the test dataset. To avoid overfitting, Kaggle partitions the test dataset into two subsets and does not

[8] https://www.kaggle.com/kaggle/meta-kaggle

Name of the Competition	Total Reward	Number of Submissions	Teams	Start Date	Deadline
Heritage Health Prize	500,000	23,421	1,221	04/04/2011	04/04/2013
Allstate Purchase Prediction Challenge	50,000	24,526	1,568	02/18/2014	05/19/2014
Higgs Boson Machine Learning Challenge	13,000	35,772	1,785	05/12/2014	09/15/2014
Acquire Valued Shoppers Challenge	30,000	25,138	952	04/10/2014	07/14/2014
Liberty Mutual Group - Fire Peril Loss Cost	25,000	14,751	634	07/08/2014	09/02/2014
Driver Telematics Analysis	30,000	36,065	1,528	12/15/2014	03/16/2015
Crowdflower Search Results Relevance	20,000	23,237	1,326	05/11/2015	07/06/2015
Caterpillar Tube Pricing	30,000	23,834	1,187	06/29/2015	08/31/2015
Liberty Mutual Group: Property Inspection Prediction	25,000	40,594	2,054	07/06/2015	08/28/2015
Coupon Purchase Prediction	50,000	18,477	1,076	07/16/2015	09/30/2015
Springleaf Marketing Response	100,000	34,861	1,914	08/14/2015	10/19/2015
Homesite Quote Conversion	20,000	28,571	1,334	11/09/2015	02/08/2016
Prudential Life Insurance Assessment	30,000	42,336	2,452	11/23/2015	02/15/2016
Santander Customer Satisfaction	60,000	93,031	5,117	03/02/2016	05/02/2016
Expedia Hotel Recommendations	25,000	22,709	1,974	04/15/2016	06/10/2016

Table 1: Summary of the Competitions in the Data (Partial List)

Note: The table only considers submissions that received a score. The total reward is measured in US dollars at the moment of the competition. See Table A.1 in the Online Appendix for the complete list of competitions.

inform participants which observations correspond to each subset. The first subset is used to generate a *public score* that is posted in real-time on a public leaderboard. The second subset is used to generate a *private score* that is never made public during the contest—it is revealed only at the end of the competition. For example, in the Heritage Health Prize, the test dataset was divided into a 30 percent subsample to compute the public scores and a 70 percent subsample to compute the private scores. Kaggle discloses the percentage of the data in each subsample, but players do not know which observations belong to each subsample, which creates imperfect correlation between public and private scores. We used competitions that used between 10 and 90 percent of the test dataset to generate public scores.

All the competitions we consider display a real-time public leaderboard which contains the public score of every submission made up to that point in time. Players' final standings, however, are calculated using the private scores, so the final standings may be different than the final standings displayed in the public leaderboard.[9] Hence, the

[9]The coefficient of correlation between public and private scores in our sample is 0.99. In about 79 percent of the competitions, the winner finished the competition within the top three of the final

public leaderboard provides informative, yet noisy, signals on the performance of all players throughout the contest.

An observation in our dataset is a submission in a contest. For each contest, we observe information on all submissions including when they were made (time of submission), who made them (team identity), and their score (public and private scores). These data allow us to reconstruct both the public and private leaderboard at every instant of time.

2.2 Motivating Facts

Our modeling choices are guided by a series of empirical facts. To make comparisons across contests, we normalize the contest length and the total prize to one, and we standardize public and private scores.

In each competition, lower scores are attributed to participants that may be not trying to win it but instead are participating for non-pecuniary motives. We are interested in modeling competitive players—those who are affected by the design of the competition and are trying to win. For this purpose, we group teams into "competitive" and "non-competitive" categories. Competitive teams are defined as teams that obtain scores above the 75th percentile of the score distribution in a competition.[10] Table 2 presents summary statistics at the competition level, team level, and submission level.

Table 2 (Panel A) presents summary statistics at the competition level. On average, there are 893.7 teams per competition, the reward is about $30,489, and competitions last for about 81.69 days. Panels B and C in Table 2 show summary statistics for all teams and competitive teams, respectively. About 25 percent of the teams are competitive and these teams send an average of 40 submissions per competition, which exceeds the overall sample average of 16.5 submissions per team. The number of members in a competitive team is on average 1.2 members, which is not significantly different than the average number of team members when considering the full sample of teams.

public leaderboard (see Table A.2 in the Online Appendix).

[10] Table A.3 in the Online Appendix shows that competitive teams are more experienced: 63 percent participate in more than one competition.

Panels D and E in Table 2 present summary statistics for all submissions and submissions by competitive teams, respectively. The standardized public and private scores are on average higher for competitive teams, but their scores also present significant variation (standard deviation of 0.75). Competitive teams also play more frequently than the rest of the teams—the average time between submissions for competitive teams and all teams is 1.2 and 1.5 percent of the contest time, respectively.

Observation 1. *Most teams have a single member.*

Figure 1 shows the evolution of the number of submissions and teams over time. Figure 1(a) partitions all the submissions into time intervals based on their submission time. The figure shows that the number of submissions increases over time, with roughly 20 percent of them being submitted when 10 percent of the contest time remains, and only 6 percent of submissions occurring when 10 percent of the contest time has elapsed. Figure 1(b) shows the timing of entry of new teams into the competition. The figure shows that the rate of entry is roughly constant over time, with about 20 percent of teams making their first submission when 20 percent of the contest time remains.

Observation 2. *New teams enter at a constant rate throughout the contest.*

We also explore the time between submissions at the team level. Figure 2 shows a local polynomial regression for the average time between submissions as a function of time. The figure shows that the average time between submissions increases over time, suggesting that either teams are experimenting when they enter the contest or that building a new submission becomes increasingly difficult over time. Combined, Figure 1 and Figure 2 suggest that the increase in submissions at the end of contests is not driven by all teams making submissions at a faster pace, but simply because there are more active teams at the end of the contest and potentially more incentives to play.

Observation 3. *The rate of arrival of submissions increases with time.*

Table 3 decomposes the variance of public scores using a regression analysis. In column 1, we find that 40 percent of the variation in public score is between-team variation, suggesting that teams differ systematically in the scores that they achieve. In column 2, we control for the number of submissions that a team has submitted up to the time

Panel A: Competition-level statistics

	N	Mean	St. Deviation	Min	Max
Number of teams	57	893.702	963.081	79	5,117
Reward quantity	57	30,488.596	66,736.377	1,000	500,000
Length (days)	57	81.69	87.90	1	700

Panel B: Overall team-level statistics

	N	Mean	St. Deviation	Min	Max
Number of submissions	50,941	16.531	29.538	1	665
Number of members	50,941	1.127	0.604	1	40
Competitive team (indicator)	50,941	0.247	0.431	0	1

Panel C: Team-level statistics — competitive teams

	N	Mean	St. Deviation	Min	Max
Number of submissions	12,591	40.078	47.904	1	665
Number of members	12,591	1.228	0.881	1	24

Panel D: Overall submission statistics

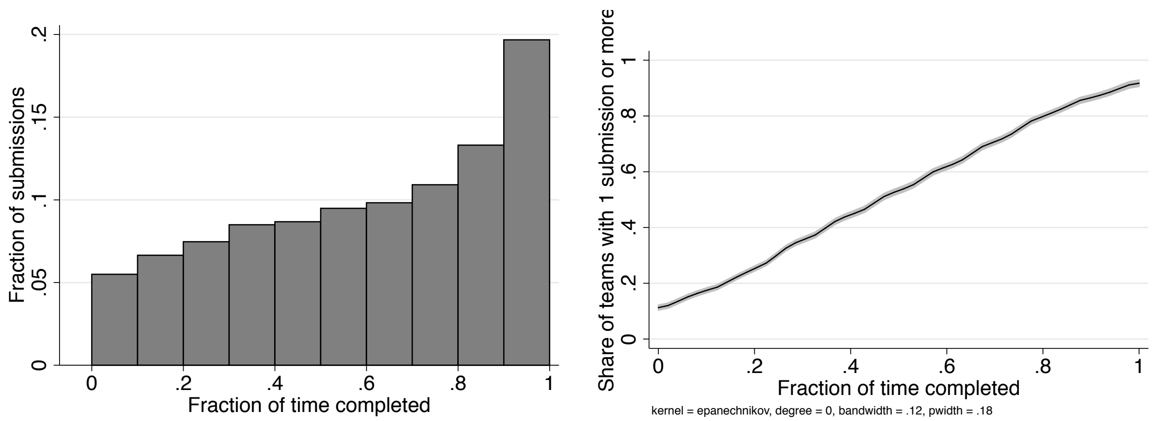
	N	Mean	St. Deviation	Min	Max
Public score	842,089	0.004	0.991	-4.000	5.659
Private score	842,089	0.005	0.991	-4.000	5.432
Time of submission	842,089	0.601	0.289	0.000	1.000
Time between submissions	791,146	0.015	0.053	0.000	0.998

Panel E: Overall submission statistics — competitive teams

	N	Mean	St. Deviation	Min	Max
Public score	504,621	0.358	0.751	-3.999	5.659
Private score	504,621	0.355	0.749	-4.000	5.432
Time of submission	504,621	0.623	0.281	0.000	1.000
Time between submissions	492,030	0.012	0.044	0.000	0.985

Table 2: Summary Statistics

Note: An observation in Panel D and E is a submission; an observation is a team–competition combination in Panels B and C; an observation in Panel A is a contest. Scores are standardized and time is rescaled to be contained in the unit interval. Time between submissions is the time between two consecutive submissions by the same team. Competitive teams are teams that achieved a public score above the 75th percentile of a contest's final distribution of scores.



(a) **(b)**

Figure 1: Submissions and Entry of Teams Over Time Across all Competitions

Note: An observation is a submission. Panel (a) shows a histogram of submission by elapsed time categories. Panel (b) shows a local polynomial regression of the number of teams with 1 or more submissions as a function of time.

	(1)	(2)	(3)	(4)
	All teams		Competitive teams	
	Public Score		Public Score	
Submission number		0.0047***		0.0041***
		(0.0000)		(0.0000)
Competition × Team FE	Yes	Yes	Yes	Yes
Observations	833,970	833,970	504,410	504,410
R^2	0.490	0.513	0.226	0.270

Table 3: Decomposing the Public Score Variance

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a submission. Submission number is defined at the competition–team–submission level and measures the number of submissions made by a team up to the time of a submission.

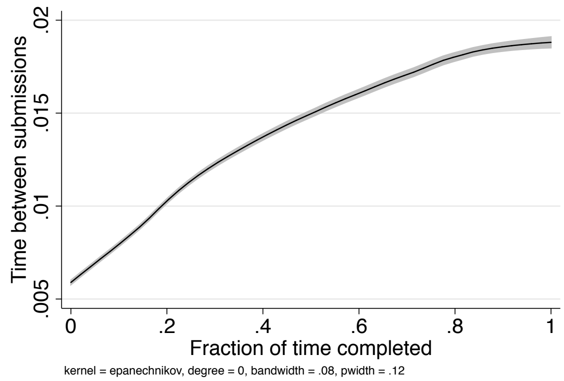


Figure 2: Time Between Submissions

Note: An observation is a submission. The figure shows a local polynomial regression of the time between submissions as a function of time.

of each submission (e.g., the variable takes the value $n-1$ for a team's nth submission). This variable allows us to capture whether learning can explain some of the variation in scores. Column 2 shows that later submissions obtain higher scores, but only an extra 2.3 percent of the variance in scores is explained by this control. This suggests that while learning may be present, between-team variation explains the majority of the systematic variation in scores. Columns 3 and 4 repeat the analysis for competitive teams. In this restricted sample, teams are more homogeneous, so team fixed-effects explains less of the variation when compared to the whole sample.

Observation 4. *Teams systematically differ in their ability to produce high scores.*

To understand how the public leaderboard shapes incentives to participate, we regress an indicator for whether a given submission was a team's last submission on the distance between the team's best public score up to that time and the best public score across all teams up to that time. Table 4 (Column 1) shows that it is more likely for teams to drop out of the competition when they start falling behind in the public score leaderboard. A one standard deviation increase in a team's deviation from the maximum public score at time t is associated with a 2.84 percent increase in the likelihood of a team dropping out of the competition at time t. Column 2 explores whether this result is heterogeneous between competitive and non-competitive teams, and shows that competitive teams are less discouraged to quit the competition when they are falling behind, compared to

	(1)	(2)
	Last submission (indicator)	
Deviation from max public score (standardized)	0.0284*** (0.000 3)	0.0138*** (0.0005)
Deviation * Competitive		-0.0156*** (0.0005)
Competitive team		-0.0646*** (0.0009)
Competition FE	Yes	Yes
Observations	842089	842089
R^2	0.020	0.040
p-value F-test		0.0000

Table 4: Indicator for Last Submission as a Function of a Team's Deviation from the Maximum Public Score

Note: Robust standard errors in parentheses.$^{*}(p < 0.1)$, $^{**}(p < 0.05)$, $^{***}(p < 0.01)$. Deviation from max public score is the difference between the maximum public score minus the score of a submission, at the time of that submission. We standardize this variable using its competition-level standard deviation. See Table 2 for the definition of competitive team.

non-competitive teams.

In Table 5, we analyze how incentives to make a new submission are affected by a submission that increases the maximum public score by a sufficient amount (e.g., 0.01 for our analysis in Table 5). We call such a submission *disruptive*. To measure how a disruptive submission affects incentives to make new submissions, we first partition time into intervals of length 0.001 and compute the number of submissions in each of these intervals. We then perform a comparison of the number of submissions before-and-after the arrival of the disruptive submission, restricting attention to periods that are within 0.05 time units of the disruptive submission. Column 1 in Table 5 shows that the number of submissions decreases immediately after a disruptive submission by an average of 2.24 percent. We take this as additional evidence of players using to the public leaderboard to make their decisions to continue or to quit. Table 5 (Column 2) shows that non-competitive teams are discouraged, in contrast to competitive teams.

	(1)	(2)	(3)	(4)
	Number of submissions (in logs)			
After disruptive submission	-0.0224***	-0.0373***	-0.0506***	-0.0395***
	(0.0081)	(0.0095)	(0.0107)	(0.0097)
After * Competitive		0.0196		
		(0.0129)		
After * Top 50			0.0586***	
			(0.0157)	
After * Top 10				0.0885***
				(0.0195)
Competition FE	Yes	Yes	Yes	Yes
Observations	21545	37666	36701	31657
R^2	0.819	0.729	0.637	0.694
p-value F-test		0.0522	0.4926	0.0045

Table 5: The Impact of Disruptive Submissions on Participation

Note: Robust standard errors in parentheses. $^{*}(p < 0.1)$, $^{**}(p < 0.05)$, $^{***}(p < 0.01)$. Disruptive submissions are those that increase the maximum public score by at least 0.01. Number of submissions is the number of submissions in time intervals of length 0.001. The regressions restrict the sample to before and after 0.05 time-units of the disruptive submission. All specifications control for time and time squared. See Table 2 for the definition of competitive team. Top 50 and Top 10 are indicators for whether the team ended the competition within the top 50 and 10 participants, respectively.

Column 3 repeats the exercise in Column 2 using instead an indicator for teams who ended in the top 50, and shows similar results. Column 4 repeats the exercise in Column 3 using instead an indicator for whether the team ended in the top 10, and shows that top 10 teams are encouraged by a disruptive submission (i.e., they increase their number of submissions after a disruptive submission). Table 5 complements Table 4 in showing that the leaderboard shapes participation incentives and the leaderboard has heterogeneous effects across players.

Observation 5. *The public leaderboard shapes participation incentives. This effect is heterogeneous across players.*

Players may strategically choose when to release a disruptive submission if they knew that a submission is disruptive. In this case, teams would have incentives to submit a

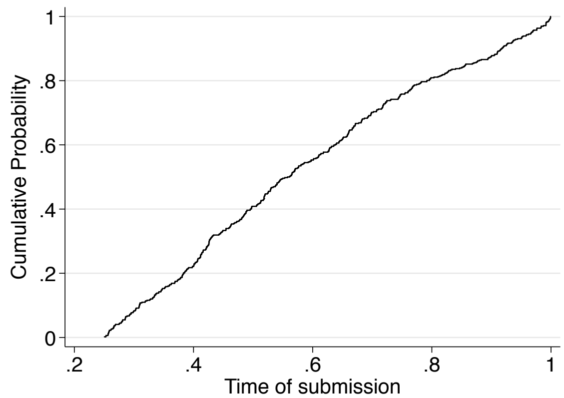


Figure 3: Timing of Drastic Changes in the Public Leaderboard's Maximum Score (i.e., Disruptive Submissions): Cumulative Probability Functions

Note: An observation is a submission that increases the maximum public score by at least 0.01. The figure plots submissions that were made when at least 25 percent of the contest time had elapsed.

disruptive submission as late as possible in the competition to avoid encouraging players who are capable of generating good scores (Column 4 in Table 5). Empirically, however, we do not find this effect. Figure 3 plots the timing of submissions that increased the maximum public score by at least 0.01. To remove outliers, in the figure we restrict attention to submissions sent after 25 percent of the contest time has elapsed. The figure suggests that disruptive submissions arrive uniformly over time and the pattern suggests that teams are either not strategic or they do not know when a submission will be disruptive. This may be driven by the fact that teams only learn about the out-of-sample performance of a submission after Kaggle has evaluated it. That is, before making a submission, the teams can only evaluate the solution using the training data, which is not fully informative about its out-of-sample performance.

Observation 6. *Submissions that disrupt the public leaderboard are submitted uniformly over time.*

3 Empirical Model

We consider each contest as a separate game where a number of players enter at a constant rate over time (Observation 2).[11] We model the time of entry of a player as a random variable, τ_{entry}, drawn from an exponential distribution of parameter $\mu > 0$. Players are heterogeneous in their abilities (Observation 4).[12] Each player is endowed with a type $\theta \in \Theta = \{\theta_1, ..., \theta_p\}$ where $\Pr(\theta = \theta_k) \equiv \kappa(\theta_k)$, $k = 1, ..., p$. We assume that a player's type is publicly known. We model the score of a submission as a random variable drawn from a type-dependent distribution $F_\theta(\cdot)$.

Players can send multiple submissions throughout the contest, but they must work on one submission at a time. The cost of building a new submission is privately observed by each player and it is independently drawn from a distribution $c \sim K$. Finishing a submission takes a player a random time τ distributed according to an exponential distribution of constant parameter λ (Observation 3). After finishing a submission, players immediately make the decision of whether to continue playing or to quit forever; this is a revision game, players have stochastic opportunities to play. Figure 4 shows the timing of the game before the end of the competition at T.

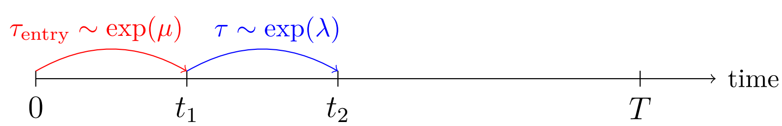


Figure 4: Timing of the game. A player enters at time t_1. At this time, the player decides to continue playing. The next submission takes time $t_2 - t_1$ to arrive. At time t_2, the player again decides to quit or play.

The collection of vectors (identity, time, score) from the beginning of the contest until instant t conforms the *public leaderboard* at time t, denoted by $\mathcal{L}_t$. In this leaderboard the scores of player i and her rivals are denoted by $y_{i,t}$ and $y_{-i,t}$, respectively. The vector of rival scores $y_{-i,t}$ transitions to $y_{-i,s}$, where $s > t$, with probability density $dG(y_{-i,s}|\mathcal{L}_t, t)$. The distribution G is an equilibrium object that is consistent with the players' equilib-

[11] We find no evidence suggesting that players strategically choose their time of entry. For instance, there is no significant correlation between the time of entry and the final ranking.

[12] We ignore team incentives and we treat each team as a single player (Observation 1).

rium strategies. The expected payoff of player i at the end of the contest, conditional on a leaderboard $\mathcal{L}_T$, is $\pi_i(\mathcal{L}_T) = \pi_i(y_{i,T}, y_{-i,T})$. Note that $\sum_i \pi_i(y_{i,T}, y_{-i,T}) = V$, where V is the total reward in the contest that we normalize to $V = 1$.[13]

Public Leaderboard. If player i (of type θ) has an opportunity to play at time $t < T$ and chooses to quit the contest, her expected payoff is

$$V^{\text{Quit}}(y_{i,t}, y_{-i,t}, t) = \int \pi_i(y_{i,t}, y_{-i,T}) dG(y_{-i,T}|\mathcal{L}_t, t).$$

This is, player i's expected payoff of quitting corresponds to finishing the contest with her current scores $y_{i,t}$ while rivals finish with scores $y_{-i,T}$. Player i computes the expected payoff over rivals' scores, under the belief that the current rivals' scores will transition from $y_{-i,t}$ to $y_{-i,T}$ with probability $dG(y_{-i,T}|\mathcal{L}_t, t)$.

If instead of quitting player i chooses to play at time $t < T$, her expected payoff is

$$V_\theta^{\text{Play}}(y_{i,t}, y_{-i,t}, t, c) = \int_t^T \int \int V_\theta(y'_{i,s}, y_{-i,s}, s) dF_\theta(y'_{i,s}) \lambda e^{-\lambda(s-t)} dG(y_{-i,s}|\mathcal{L}_t, t) dy'_{i,s} ds$$
$$+ e^{-\lambda(T-t)} V^{\text{Quit}}(y_{i,t}, y_{-i,t}, t) - c.$$

The first term in the right-hand side of the expression above corresponds to the event where the new score arrives at time $s \in (t, T)$, which happens with likelihood $\lambda e^{-\lambda(s-t)}$. In this case, the scores of player i's rivals will have transitioned from $y_{-i,t}$ to $y_{-i,s}$ with likelihood $dG(y_{-i,s}|\mathcal{L}_t, t)$, and player i's scores to $y'_{i,s}$ with likelihood $dF_\theta(y'_{i,s})$.[14] Immediately upon the arrival of the new submission at time s, and conditional on the leaderboard at time s, player i decides to continue playing or to quit. The expected payoff of this decision is captured by the value function $V_\theta(y'_{i,s}, y_{-i,s}, s)$. In the complementary event that a new submission does not arrive before the end of the contest, which occurs with probability $e^{-\lambda(T-t)}$, player i's payoff of the same as if she would have quit at time t. Finally, the cost of a new submission is c.

[13] The only effect of this normalization is that the distribution of costs $c \sim K$ must be interpreted as costs that are proportional to the size of the prize.

[14] With a slight abuse of notation, $dF_\theta(y'_{i,s})$ is the likelihood that player i's new vector of scores at time s is $y'_{i,s}$. This new vector of scores is constructed by adding a new score drawn from the distribution $F_\theta(\cdot)$ to player i's current the vector of scores $y_{i,t}$. For instance, if only the maximum score is payoff-relevant, $y'_{i,s} = \max\{z, y_{i,t}\}$ where $z \sim F_\theta(\cdot)$.

Player i's chooses to play if and only if $V^{\text{Quit}}(y_{i,t}, y_{-i,t}, t) \leq V_\theta^{\text{Play}}(y_{i,t}, y_{-i,t}, t, c)$, i.e.,

$$\begin{aligned} c \leq \int_t^T \int \int & V_\theta(y'_{i,s}, y_{-i,s}, s) dF_\theta(y'_{i,s}) \lambda e^{-\lambda(s-t)} dG(y_{-i,s}|\mathcal{L}_t, t) dy'_{i,s} ds \\ & - (1 - e^{-\lambda(T-t)}) V^{\text{Quit}}(y_{i,t}, y_{-i,t}, t), \end{aligned} \tag{1}$$

where the value function is

$$V_\theta(y_{i,t}, y_{-i,t}, t) = \mathbb{E}_c \left[\max\{V_\theta^{\text{Play}}(y_{i,t}, y_{-i,t}, t, c), V^{\text{Quit}}(y_{i,t}, y_{-i,t}, t)\} \right].$$

The equilibrium distribution G captures dynamic participation effects: the score of a player's submission disclosed on the leaderboard affects future participation decisions by their rivals. It is noteworthy to mention that Equation 1 captures a "discouragement effect": higher rival scores decrease the right-hand side of the inequality.

The equilibrium concept we use is Markov perfect equilibrium. The decision to play or to quit for any player facing leaderboard $\mathcal{L}_t$, at time t, and the submission cost c, is the same. Thus, the (pure) equilibrium strategy of player i, conditional on ($\mathcal{L}_t$, t, c) is to play iff Equation 1 holds. The existence of an equilibrium follows from arguments analogous to those in Aguirregabiria and Mira (2007).

No Public Leaderboard. Without a public leaderboard, only the realization of each player's own scores can be used to condition the decision of continuing to play or quitting. The decision problem is similar to a model of sequential search where players form beliefs about the distribution of final scores. In a similar setting, Taylor (1995) shows that players' optimal strategy is a stopping rule: players will stop when their own score is higher than a threshold. Let $W_{-i}(\cdot)$ be the distribution of *final scores* on the leaderboard for all players except player i. This is an equilibrium object capturing the strategies of the rivals of player i. If player i's own scores at time t are $y_{i,t}$ and she decides to quit the contest, her expected payoff is

$$V^{\text{Quit}}(y_{i,t}) = \int \pi_i(y_{i,t}, y_{-i}) dW_{-i}(y_{-i}),$$

where y_{-i} are the scores of player i's rivals at the end of the contest, which are distributed according to the (time-independent) distribution $W_{-i}(\cdot)$. If instead of quitting the contest player i chooses to play, her expected payoff is

$$V_\theta^{\text{Play}}(y_{i,t}, t, c) = \int_t^T \int V_\theta(y'_{i,s}, s) dF_\theta(y'_{i,s}) \lambda e^{-\lambda(s-t)} ds + e^{-\lambda(T-t)} V^{\text{Quit}}(y_{i,t}) - c.$$

In the expression above, in the event of a new score arriving at time $s \in (t, T)$, which happens with likelihood $\lambda e^{-\lambda(s-t)}$, player i's new scores are $y'_{i,s}$ with likelihood $dF_\theta(y'_{i,s})$. At time s, player i decides to either continue playing or to quit. The expected value of this decision is captured by the value function $V_\theta(y'_{i,s}, s)$. In the complementary event, with probability $e^{-\lambda(T-t)}$ the new submission does not arrive before the end of the contest, and player i's payoff is the same as if she would have quit at time t.

Player i's optimal decision is to play if and only if

$$c \leq \int_t^T \int V_\theta(y_{i,s}, s) f_\theta(y'_{i,s}) \lambda e^{-\lambda(s-t)} ds - (1 - e^{-\lambda(T-t)}) V^{\text{Quit}}(y_{i,t}) \tag{2}$$

where the value function is

$$V_\theta(y_i, t) = E_c \left[\max\{V_\theta^{\text{Play}}(y_{i,t}, t, c), V^{\text{Quit}}(y_{i,t})\} \right].$$

Note that each distribution $W_{-i}(\cdot)$ (uniquely) determines the function $V^{\text{Quit}}(y_{i,t})$, the expected payoff of quitting conditional on scores $y_{i,t}$.

We also use the equilibrium concept of Markov perfect equilibrium. However, the decision to play or to quit for any player depends only on her current scores, the current time, and submission cost. Thus, the (pure) equilibrium strategy of player i, conditional on (y_{it}, t, c) is to play iff Equation 2 holds. The existence of an equilibrium again follows from arguments analogous to those in Aguirregabiria and Mira (2007).

3.1 Discussion of Modeling Assumptions

Some of the assumptions in our model are made for computational tractability or to keep the model parsimonious, whereas others are justified from empirical observations.

Our analysis does not incorporate learning both because of tractability and because Table 3 shows that between-team differences explain the majority of the systematic variation in scores. Some teams may experiment and improve their performance over time, but we show in Table 3 that this effect is of second order relative to the variance in scores that can be explained by the "innate" ability of players.[15]

[15] Clark and Nilssen (2013), for example, present a theory of learning by doing in contests.

A second assumption of our model is that entry is exogenous. In reality, players choose which contests to participate in. Azmat and Möller (2009) show that contest design (in particular, the allocation of prizes) affects players decisions when they choose among multiple contests. Levin and Smith (1994), Bajari and Hortacsu (2003), and Krasnokutskaya and Seim (2011) explore how endogenous entry affects equilibrium outcomes and optimal design in auctions. Although we acknowledge this shortcoming of our analysis, we have various reasons to make this assumption. First, in our data most players participate in a single contest (see Table A.3 in the Online Appendix), so it is hard to define a group of potential entrants. Second, all contests in Kaggle display a leaderboard, so we cannot identify how this feature of contest design (displaying a leaderboard) affects entry using the observational data. Finally, and as we will discuss below, our experiment reveals that contests with and without a public leaderboard draw the same amount of participants on average, which alleviates the concern of endogeneous entry.

A third potential concern is the assumption that players do not strategically choose when to send their submissions. Ding and Wolfstetter (2011) show that players could withhold their best solutions and negotiate with the sponsor of the contest after the contest has ended. This selection introduces a bias on the quality of submitted solutions. In our setting, players benefit by sending a submission, because they receive a noisy signal about the performance of the submission. We also find that the timing of disruptive submissions is roughly uniformly distributed over time (as shown in Figure 3) and that there is no correlation between the final ranking and the time of entry, which alleviates the concern about strategic timing of submissions.

Also related is the assumption that players make a decision to continue or to quit immediately after the arrival of a submission. If we observe two submissions by a player at times t_1 and t_2, we know that this player must have spent some time $t \in [0, t_2 - t_1]$ working on the submission. Instead of modeling the distribution of idle time between submissions—similar to the random time of play assumption in Arcidiacono et al. (2016)—we assume that the idle time is zero, i.e, $t = t_2 - t_1$. We make this assumption because we observe a short time between submissions. Thus, the effect of idle time is likely to be small, but adding this effect would incorporate an extra burden in the estimation of our model.

4 Estimation

For each contest in our dataset we estimate the parameters of the model in two steps. We first estimate a number of primitives without using the full structure of the model. We then use these estimates for the estimation of the remaining parameters using a likelihood function based on the model. When estimating the model, we restrict attention to the subsample of competitive teams (see Table 2).

The full set of parameters for a given contest include: i) the distribution of new player arrival times, which we assume follow an exponential distribution with parameter μ; ii) the distribution of submission arrival times, which we assume follow an exponential distribution with parameter λ; iii) the distribution of private score conditional on public score, $H(\cdot|p^{public})$, which we assume is given by $p^{private} = \alpha + \beta p^{public} + \epsilon$, with ϵ distributed according to a double exponential distribution; iv) the type-specific cumulative distribution of public scores, which we assume is given by the standard normal distribution, $Q_j(x) = \Phi\left(\frac{x-\theta_j^{mean}}{\theta_j^{st.dev}}\right)$ for type θ_j; v) the distribution of types, κ, which we assume is a discrete distribution over the set of player types, Θ; and, lastly, vi) the distribution of submission costs, which we assume has a support that is bounded above by 1 (i.e., the normalized value of the total prize money), and has a cumulative distribution function given by $K(c;\sigma) = c^{\sigma}$ (with $\sigma > 0$).

We estimate primitives i) through v) in a first step that does not require the full structure of the model, and vi) using the likelihood function implied by the model. i), ii), and iii) are estimated using the maximum likelihood estimators for μ, λ, and (α, β), respectively. iv) and v) are specified as a Gaussian mixture model that we estimate using the EM algorithm. The EM algorithm estimates the k Gaussian distributions (and their weights, $\kappa(\theta_k)$) that best predict the observed distribution of public scores. Throughout our empirical analysis we assume that there are $k = 2$ player types.[16] Section B in the Online Appendix provides additional details of the estimation procedure.

We estimate the distribution of costs using the likelihood function implied by the model, which is based on the decision of a player to make a new submission. Recall that a player can make a new submission immediately after the arrival of her previous submission.

[16]We experimented with different number of types. $k = 2$ is parsimonious and gave us a good fit.

A player of type θ facing state variables s chooses to make a new submission at time t if and only if

$$c \leq \Gamma_{\theta,t}(s), \tag{3}$$

where c is the cost of a submission and $\Gamma_{\theta,t}(s)$ are the net benefits of making a new submission at time t for a player of type θ given state variables s. This is, $\Gamma_{\theta,t}(s)$ corresponds to the right hand-side of Equation 1 in competitions with a leaderboard. $\Gamma_{\theta,t}(s)$ depends on primitives estimated in the first step of the estimation, player type, and the equilibrium conditional choice probabilities (CCPs) reflecting the equilibrium behavior of all players in the remaining time of the contest. Based on Equation 3, a θ-type player facing state variables s plays at time t with probability

$$\Pr(\text{play}|s,t,\theta) = \Pr(c \leq \Gamma_{\theta,t}(s)) = K\left(\Gamma_{\theta,t}(s);\sigma\right).$$

The likelihood function is constructed using tuples $\{(s_i, t_i, t_i', \theta_i)\}_{i\in N}$, where i is a submission, s_i is the vector of state variables at the moment of making the submission, t_i is the submission time, t_i' is the arrival time of the next submission, and θ_i is player type. Because t_i' is not recorded for the last observed submission of each player, we simply set t_i' to $t_i' > T$ for the last observed submission of each player, and address this censoring when constructing the likelihood function. The type of each player is not observed in the data, but we use the vector of scores of each player combined with estimates both of the distribution of player types and the type-specific distribution of scores to compute the posterior probability of a player being of type θ_j. We then use the Bayes classifier to assign a type to each player (see Section B in the Online Appendix for details). We assume that our estimates for the type of each player matches the information held by the players in the game.

The likelihood function considers two cases. If the next submission arrives before the end of the contest, i.e., $t_i < t_i' \leq T$, then the player must have chosen to make a new submission at t_i, and the likelihood of the observation $(s_i, t_i, t_i', \theta_i)$ is given by $l(s_i, t_i, t_i', \theta_i) = \Pr(\text{play}|s_i, t_i, \theta_i) \cdot \lambda e^{(-\lambda(t_i'-t_i))}$, where $\lambda e^{(-\lambda(t_i'-t_i))}$ is the density of the submission arrival time. If the submission at time t_i was the player's last recorded submission, i.e., $t_i' > T$, then the likelihood of $(s_i, t_i, t_i' > T, \theta_i)$ is given by $l(s_i, t_i, t_i' > T, \theta_i) = \Pr(\text{play}|s_i, t_i, \theta_i) \cdot e^{(-\lambda(T-t_i))} + 1 - \Pr(\text{play}|s_i, t_i, \theta_i)$, which considers both the events of i) the player choosing to make a new submission at t_i and the submission arriving after the end of the contest; and ii) the event of the player choosing not to make a new submission.

Evaluating the likelihood function requires computing the net benefits of making a new submission for a player of type θ at time t and given state variables s (i.e., $\Gamma_{\theta,t}(s)$), which depend on the equilibrium CCPs. A full-solution method where we compute the equilibrium of the game for every trial vector of parameters is computationally infeasible given the dimensionality of the state space. Hence, we use a CCP-based estimator, which makes use of estimates of the equilibrium CCPs to simulate the net benefits of making a new submission (Hotz and Miller, 1993; Hotz et al., 1994). We simulate the net benefits of making a new submission (as opposed to computing them based on the value functions) mainly because of the dimensionality of the state space, but also because it saves us from having to discretize the state space, which is costly because the scores on the leaderboard are continuous variables. We provide details on how we estimate the CCPs and how we simulate the net benefits of making a new submission in Section B in the Online Appendix.

While the CCP-based estimator makes the estimation problem feasible, it is still computationally costly because the econometrician needs to simulate continuation histories of the game for every observation in the estimation sample in games with hundreds or even thousands of players playing repeatedly.[17] As a consequence, we estimate the parameters of the cost distribution of each contest on a (stratified) random sample of 100 observations, and make use of the weighted maximum likelihood estimator in Manski and Lerman (1977), where the weights are used to correct for the oversampling (or undersampling) of each strata.[18]

[17]In the estimation, we simulate the net benefits of making a new submission for every observation in the estimation sample based on 200 simulations of continuation histories.

[18]The stratified random sample of each contest includes four stratas. For every type of player, the sample includes i) submissions that are the last recorded submission of a player, and ii) submissions that are not the last recorded submission of a player.

4.1 Model Estimates

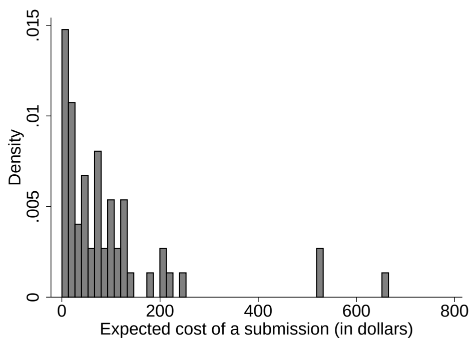
Table 6 presents the maximum likelihood estimates for the submission-cost distribution as well as for the distributions of entry time and submission arrival time. Table A.5 in the Online Appendix presents the EM algorithm estimates for the type-specific distributions of scores, and Table A.6 in the Online Appendix presents estimates for the distribution of private scores conditional on public scores. All primitives of the model were estimated separately for each contest.

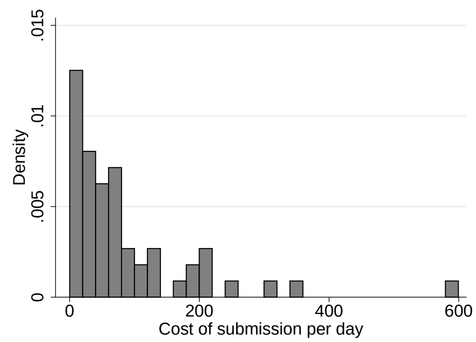
Table 6 (Column 1) shows estimates for the players' rate of entry in a given competition. The estimates imply that the average entry time ($1/\mu$) ranges between 22 and 63 percent of the contest time, and the mean average entry time across all contests is 41 percent of the contest time. Table 6 (Column 3) presents the estimates for the rate at which submissions are completed. In line with Table 2, the estimates suggest that the average time between submissions ($1/\lambda$) ranges between 0.5 and 5.5 percent of the contest time, and the mean average time between submissions across all contests is 1.5 percent of the contest time.

Table 6 (Column 5) presents estimates for the coefficients governing the distribution of submission costs. These estimates imply that the expected submission cost ranges between 2.2 and 531 dollars (percentiles 5 and 95, respectively).[19] Figure 5 shows some implications of our estimates. Figure 5(a) shows the distribution of the expected cost of making a submission (in dollars), and Figure 5(b) shows the daily cost of working on a submission (in dollars). The average values for the expected cost of a submission and the daily cost of a submission are 170.77 and 99.78 dollars, respectively. Figure 5(c) shows a scatter plot of the total expected cost spent by all participants of a contest and the prize, both measured in logs. We can see that in the majority of the contests the total expected cost is greater than the prize. This is a feature of rent-seeking contests: competition pushes participants to exert inefficiently high levels of effort relative to the size of the prize (rent dissipation).

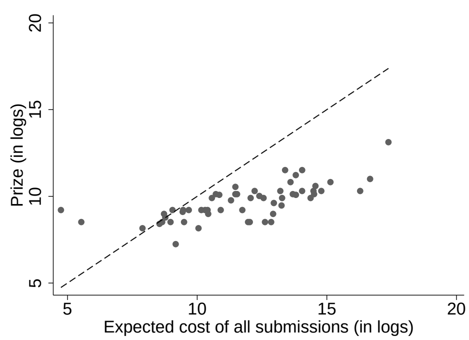
Table 7 studies how the estimates for the average cost of making a submission, the rate of team entry, and the rate of arrival of submissions vary as a function of contest prize.

[19] The expected cost in dollars is given by $E[c] = E\left[\frac{c}{V}\right] V = \frac{\sigma V}{1+\sigma}$, where V is the total reward.





(a) Cost of a submission (in dollars) **(b)** Cost of a submission per day (in dollars)



(c) Rent dissipation

Figure 5: Estimates for the cost of making a submission

Note: An observation is a contest. Cost of a submission per day is the expected cost divided by the average number of days between submissions. The average values for the expected cost of a submission and the daily cost of a submission are 170.77 and 99.78 dollars, respectively. The expected cost of all submissions is the expected cost of a submission multiplied by the predicted number of submissions for each contest. The predicted number of submissions is based on 200 simulations of each contest using our model estimates.

	μ	SE	λ	SE	σ	SE	$\log L(\hat{\delta})/N$	N
hhp	2.585	0.1701	191.9182	1.6088	0.0089	0.0004	-3.809	14231
allstate-purchase-prediction-challenge	1.9856	0.1276	125.4499	1.1689	0.0009	0.0001	-2.6272	11519
higgs-boson	2.3698	0.114	122.1003	0.8177	0.0043	0.0002	-3.6653	22298
acquire-valued-shoppers-challenge	2.0772	0.1316	165.2723	1.2866	0.0032	0.0002	-3.846	16500
liberty-mutual-fire-peril	3.3432	0.3202	122.5353	1.3388	0.001	0.0001	-3.7003	8377
axa-driver-telematics-analysis	2.4434	0.1408	127.4859	0.8925	0.0003	0.0001	-3.3961	20405
crowdflower-search-relevance	2.2697	0.1163	79.1331	0.6272	0.0047	0.0003	-3.193	15919
caterpillar-tube-pricing	3.2701	0.1758	68.2329	0.5562	0.0025	0.0002	-3.0241	15047
liberty-mutual-group-property-inspection-prediction	3.1055	0.1152	67.0227	0.4112	0.0047	0.0002	-2.8525	26573
coupon-purchase-prediction	2.0586	0.1093	73.1853	0.6539	0.005	0.0003	-1.5545	12526
springleaf-marketing-response	3.0405	0.1689	97.3279	0.7153	0.0014	0.0001	-2.9404	18513
homesite-quote-conversion	2.4958	0.1548	128.5381	0.9678	0.0035	0.0002	-3.3492	17638
prudential-life-insurance-assessment	2.16	0.0799	78.0741	0.4707	0.0068	0.0003	-2.7802	27512
santander-customer-satisfaction	2.3098	0.0563	75.1579	0.3048	0.0037	0.0002	-3.108	60816
expedia-hotel-recommendations	2.2792	0.086	43.2208	0.3422	0.0042	0.0003	-1.7432	15948

Table 6: Maximum Likelihood Estimates of the Cost and Arrival Distributions (partial list). Note: The model is estimated separately for each contest. Asymptotic standard errors are reported in the columns that are labeled 'SE.' See Table A.4 in the Online Appendix for the full table.

Table 7 (Columns 1 and 2) show a positive correlation between the contest reward and both the average cost of making a submission and the rate of new submissions, which suggests that contests with larger prizes are more difficult and participants send submissions more frequently. The greater difficulty is consistent with the empirical observation that teams remain active for less time in competitions with greater rewards (i.e., the exit rate is higher). To capture this pattern in the data, the model needs a larger cost in order to fit the larger exit rate.[20] Table 7 (Column 3) shows that entry rates are not significantly correlated with the size of the prize. This is further evidence that the timing of entry is likely not strategically chosen. Finally, Figure A.1 in the Online Appendix presents a scatter plot of the entry rate of teams and the arrival rate of submissions, and shows a weak negative correlation.

With respect to how well the model fits the data, Figure 6 plots the actual versus the predicted number of submissions in each contest. The predicted number of submissions in a contest is computed by averaging the number of submissions across 200 simulations of each contest. The simulations make use of the estimates of the model and take the number of teams that participate in each contest as given. Figure 6(a) shows the model

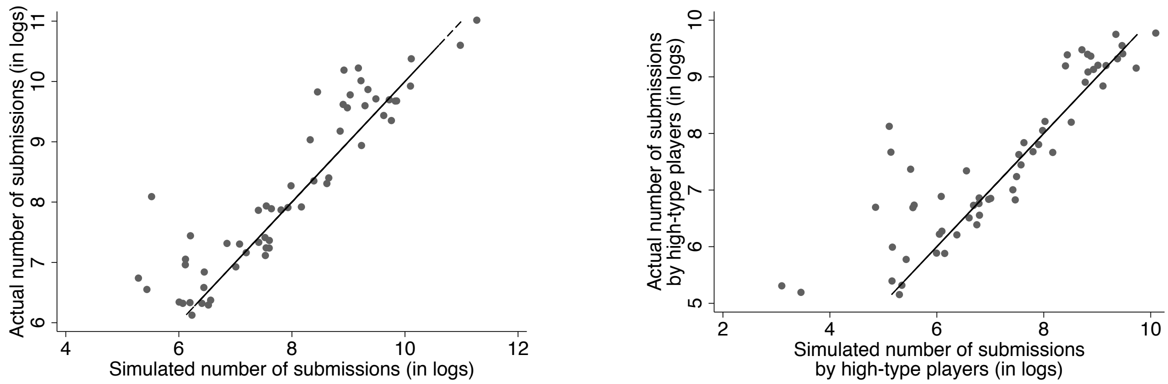
[20]Almost every competition in our data lasts three months, so the data offers little variation in contest length to establish relationships between parameter estimates and contest length.

	(1)	(2)	(3)
	$\log E[c]$	$\log \lambda$	$\log \mu$
log Prize (in USD)	1.073***	0.209***	-0.004
	(0.136)	(0.039)	(0.022)
Observations	57	57	57
R^2	0.411	0.236	0.000

Table 7: Parameter estimates and contest observables

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a contest. Expected cost is measured in dollars and is given by $E[c] = E[c/V]V = \sigma/(1+\sigma)V$, where V is the total prize. The values of σ, μ, λ are reported in Table A.4.

fit in terms of total number of submissions. The figure shows that the model does not systematically over- or under-predict participation. The correlation between the actual and the predicted number of submissions is 0.92. Figure 6(b) shows the model fit in terms of total number of submissions by high-type players, where type θ_i is the high type if $\theta_i^{mean} + 3\theta_i^{st.dev} > \theta_j^{mean} + 3\theta_j^{st.dev}$. In this case, the correlation between the actual and the predicted number of submissions is 0.88. Figure 7 shows the fit of the EM algorithm estimates for the score distributions of two competitions. These figures combined suggest a good fit of the model to the data along the dimensions of total number of submissions, total number of submissions by high types, and type-dependent distribution scores.

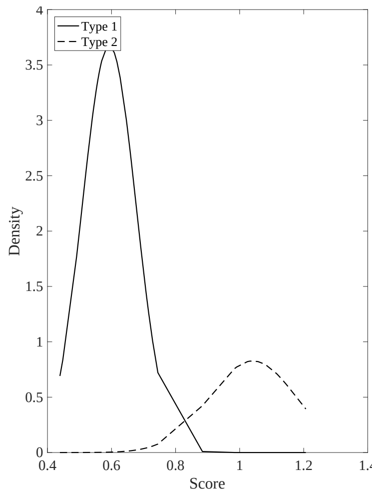


(a) Number of submissions **(b)** Number of submissions by high-type players

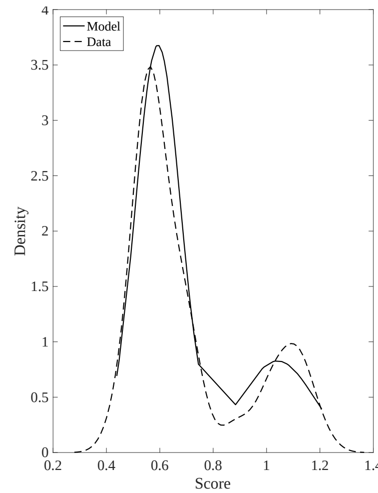
Figure 6: Model prediction for various contest outcomes

Note: An observation is a contest. The solid line is the 45-degree line. The coefficient of correlation between the actual and predicted number of submissions is 0.92; and between the actual and predicted number of submissions by high-type players is 0.88. Estimates of the type-specific score distributions are used to compute posterior probabilities of the type of each player. Players are then classified between low and high type using the Bayes classifier (see Section B in the Online Appendix for details). The predicted outcomes are based on 200 simulations of each contest using our model estimates.

Contest: ChessRatings2

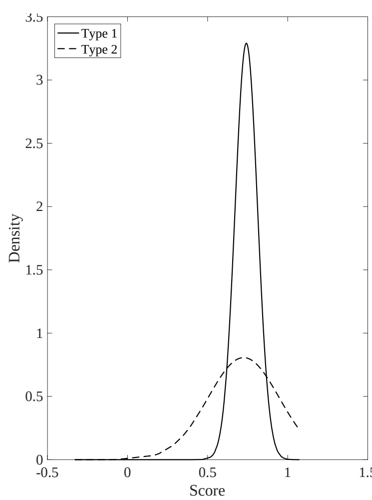


(a) Estimated distribution of scores by type.

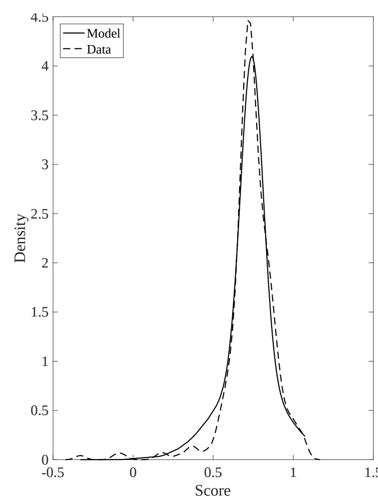


(b) Empirical distribution of scores and model fit.

Contest: DontGetKicked



(c) Estimated distribution of scores by type.



(d) Empirical distribution of scores and model fit.

Figure 7: Estimates of the distribution of scores by type for two contests

Note: EM algorithm estimates for the distribution of scores by type are reported in Table A.5 in the Online Appendix.

5 Counterfactual Information Design

We investigate the impact of information disclosure, in the form of a real-time public leaderboard, on contest outcomes. We do this by comparing contest outcomes in the equilibria with and without a public leaderboard. We compare the total number of submissions and the maximum score in these two cases. The total number of submissions is a proxy for diversity whereas the maximum score measures the quality of the "best solution." To do this, we recompute the equilibrium of the game in a counterfactual design where the sponsor does not display the public leaderboard: participants only observe their own scores but they do not observe their rivals' scores.[21]

How does performance feedback in a real-time public leaderboard impact players' incentives? Consider first the case of a contest with a public leaderboard. In this case, a player who has the opportunity to play at a given time t observes her current ranking before deciding to continue playing or to quit. A history with a high maximum score on the leaderboard discourages players to continue playing. If the history of the contest was exactly the same, but players did not have access to the leaderboard, a player with low scores may choose to continue playing. The reason is that without a public leaderboard players follow a stopping-rule, and players with low scores keep playing if they have not reached the stopping threshold. On the other hand, a player with no information may stop playing too soon relative to a player who has access to a public leaderboard. For instance, a player without information could stop early in the contest after drawing a score above the threshold dictated by her stopping rule. Had this player observed the scores on the leaderboard, she would have realized that her chances of winning were slim giving her current standing in the competition. Thus, she may have continued drawing scores beyond her stopping-rule threshold, dictated by the equilibrium without a leaderboard, to improve her changes of winning.

Table 8 reports estimates for the change in (average) contest outcomes when comparing the cases of a contest without and with a public leaderboard. Table 8 (Column 1) shows that hiding the public leaderboard on average reduces the total number of submission

[21] Taylor (1995) focuses in the unique symmetric equilibrium. Our numerical analysis suggests the existence of one symmetric equilibrium, where all players of a type use the same strategy, so we restrict attention to this equilibrium.

	(1)	(2)	(3)	(4)
	Number of submissions (in logs)			
	All players	Low-type players	High-type players	Max score (in logs)
No leaderboard	-0.210**	0.205	-0.237*	-0.017**
	(0.088)	(0.159)	(0.121)	(0.007)
Observations	114	114	114	114
R^2	0.944	0.873	0.903	0.996

Table 8: The impact of the leaderboard on contest outcomes

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a contest–design combination. All specifications include contest fixed effects. The outcome variables in Columns 1-3 are the number of submissions by all players, the number of submissions by low-type players, and the number of submissions high-type players (all in logs), respectively. Type θ_i is defined as the high-type if $\theta_i^{mean} + 3\theta_i^{st.dev} > \theta_j^{mean} + 3\theta_j^{st.dev}$.

by 21 percent. Table 8 (Columns 2 and 3) show that this reduction is driven by fewer submissions by high-type players, where we define type θ_i to be the high type if $\theta_i^{mean} + 3\theta_i^{st.dev} > \theta_j^{mean} + 3\theta_j^{st.dev}$ (see Table A.5 for the type-specific parameter estimates). Table 8 (Column 4) shows that the maximum score on average decreases by 1.7 percent, which is a consequence of the decrease in the number of submission by high-type players.

To study the impact of information disclosure in more detail, we examine the total number of submissions of a player in three different scenarios. First, we compute the total number of submissions of a player in a contest without a leaderboard ($n_{no\ lb}$). Second, we compute the total number of submissions of a player in a contest with a leaderboard (n_{lb}). And third, we compute the total number of submissions of a player in a contest where everyone has access to a leaderboard but her ($n'_{no\ lb}$). We then decompose the difference in the number of submissions with and without a leaderboard into two terms:

$$\underbrace{n_{lb} - n_{no\ lb}}_{\text{Total Effect}} = \underbrace{(n_{lb} - n'_{no\ lb})}_{\text{Information Disclosure Effect}} + \underbrace{(n'_{no\ lb} - n_{no\ lb})}_{\text{Equilibrium Adjustment}} \tag{4}$$

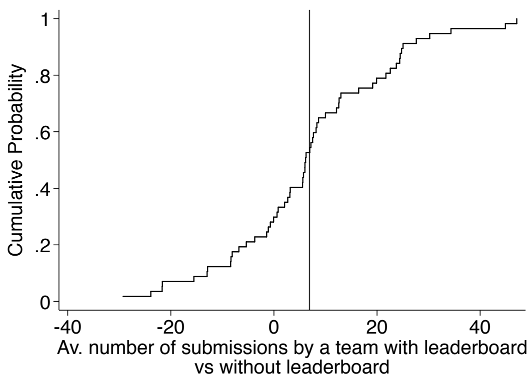
The first term in Equation 4, the information disclosure effect, captures the increase/decrease in the number of submissions of a player with access to a public leaderboard relative to the case where the same player does not have access to a leaderboard, under the assumption that all her rivals behave according to the equilibrium strategy of a contest with a leaderboard. The second term in Equation 4, the equilibrium adjustment effect, captures the change in the number of submissions by a player who does not observe the leaderboard when facing rivals who play according to the equilibrium strategy of a competition with a leaderboard ($n'_{no\ lb}$) relative to the case of facing rivals who play according to the equilibrium strategy of a competition without a leaderboard ($n_{no\ lb}$).

Figure A.2 shows that the total effect of the leaderboard on the number of submissions is mostly driven by the information disclosure effect. That is, players' strategies are different with and without a leaderboard, but the main difference is the possibility of conditioning the decision to quit on the leaderboard's information, rather than on different beliefs about the strategies used by rival players. Note that the information disclosure effect captures the difference in participation when a player experiences discouragement from lagging behind in the competition (the case with a leaderboard) relative to the case of a player considering whether to cannibalize her past scores while making a probabilistic assessment of her current position in the contest (the case without a leaderboard). Thus, the information disclosure effect is one way of measuring the relative weight of these effects on the incentives to play.

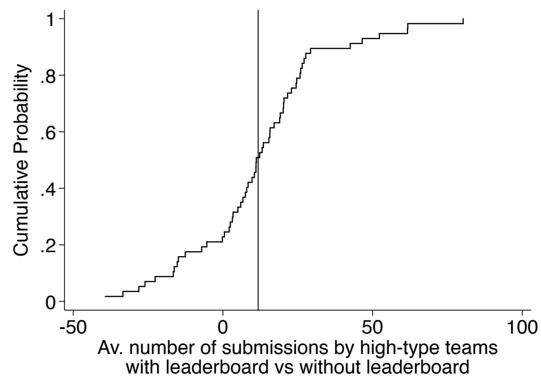
Table 8 shows that hiding a public leaderboard on average worsens contest outcomes. However, in line with theoretical results on information disclosure, a public leaderboard does not always improve contest outcomes (Rieck, 2010). Figure 8 shows the heterogeneity in the direction and magnitude of the change in contest outcomes when comparing the cases with and without a public leaderboard. The figure shows that, for most of the contests, displaying a public leaderboard improves contest outcomes. For instance, Figure 8(c) shows that the maximum score decreases with a public leaderboard only for about 20 percent of the contests.

Lastly, we test some theoretical predictions regarding the heterogeneous effects of a leaderboard on contest outcomes. Rieck (2010) shows that a leaderboard may increase or decrease contest outcomes depending on the cost-to-prize ratio and the shape of the distribution of scores.[22] We regress the total effect and the information disclosure

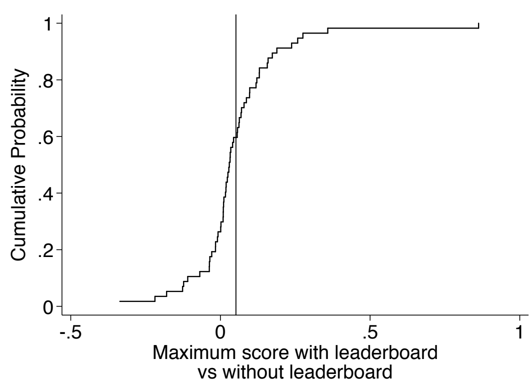
[22]Specifically, Rieck (2010) provides a condition for the public leaderbord to increase the expected



(a) Average number of submissions



(b) Average number of submissions by high-type teams



(c) Maximum score

Figure 8: The impact of the leaderboard on contest outcomes: Heterogeneity analysis

Note: An observation is a contest. Each figure plots the difference between the equilibrium outcome with a leaderboard and the equilibrium outcome without the leaderboard. Equilibrium outcomes are based on 200 simulations of each contest using our model estimates. The vertical line in each plot shows the expected value of the plotted distribution.

effect in Equation 4 on the expected cost-to-prize ratio and on the variance of the distribution of scores of high-type players. Table 9 shows that the difference in the average number of submissions by a team with a leaderboard and without a leaderboard increases both with a higher cost-to-prize ratio and a higher variance of the distribution of scores of high-type players. The same relationship holds true for our measure of the value of information.[23] Intuitively, information is more valuable when the cost of a new submission is higher. Information is irrelevant in the extreme case that new submissions are free—it is strictly dominant to make new submissions as often as possible. When the cost of new submissions is high, the cost of erring on the side of staying in the contest for too long is also higher. Regarding the variance of the distribution of scores of high-type players, a higher variance reduces the value of knowing who is leading the competition, as competition leaders change more frequently. At the same time, it encourages players with low scores to keep drawing, so the stopping threshold increases. We show that these effects increase participation on average.

6 Experimental Evidence

To complement our structural estimates, we ran a randomized control trial on Kaggle.[24] The objective of the experiment is to provide additional evidence—independent of our model's assumptions—on how information disclosure impacts participation and contest outcomes. The experiment allowed us to observe contest outcomes in competitions with and without a leaderboard, keeping other aspects of the contest fixed (e.g., difficulty, prize, duration, number of participants).

maximum score.

[23] As a robustness check, Table A.7 in the Appendix expands the results in Table 9 by including more covariates. None of the coefficients on these additional covariates are significant at a 5-percent level of significance.

[24] Approval from the University of Illinois Human Subjects Committee, IRB18644.

	(1)	(2)
	Total effect ($n_{lb} - n_{no\ lb}$)	Information Disclosure ($n_{lb} - n'_{no\ lb}$)
σ	3.7055**	4.5542***
	(1.4468)	(1.4690)
$\theta^{st.dev}_{high\ type}$	4.8077***	4.6962***
	(1.5236)	(1.5545)
Observations	57	57
R^2	0.171	0.190

Table 9: Explaining the impact of the leaderboard on contest outcomes

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a contest. Total effect is the difference in the average number of submissions by a team with a leaderboard and without a leaderboard. Information disclosure is the difference between the average number of submissions by a team in the equilibrium with a leaderboard and the average number of submissions by the same team when this team that does not observe the leaderboard but knows that all other teams observe the leaderboard. All independent variables are measured in standard deviations. σ is the parameter of the cost distribution; $\theta^{st.dev}_{high\ type}$ is the standard deviation of the distribution of scores of high-type players. Type θ_i is defined as the high-type if $\theta_i^{mean} + 3\theta_i^{st.dev} > \theta_j^{mean} + 3\theta_j^{st.dev}$.

6.1 Description of the Experiment

We hosted 44 competitions on Kaggle and each competition was randomly assigned to the treatment or control groups. Our treatment competitions displayed a real-time leaderboard, providing information about the performance of all participants, whereas our control competitions did not provide feedback to players. All of the competitions were identical in other aspects of design. The competitions were run simultaneously and lasted for 10 days. The competitions entailed solving a simple prediction problem: to interpolate a function (see Online Appendix C for details). Participants were allowed to submit up to 10 sets of predictions per day. The most accurate predictions in each competition were awarded an Amazon gift card worth \$50.

We recruited 220 students (both undergraduates and graduates) from the University of Illinois at Urbana-Champaign, via emails, department newsletters, and flyers. Participants were asked to complete an initial survey from which we obtained information

	Invited players			Entrants		
Variable	Control	Treatment	t-stat	Control	Treatment	t-stat
participated_past	0.236	0.191	-0.954	0.202	0.244	0.507
software_code	0.964	0.973	0.403	0.968	0.987	0.909
stat_tools	0.882	0.836	-0.883	0.887	0.934	0.822
mach_learning	0.536	0.518	-0.276	0.615	0.607	-0.082
regression	0.736	0.709	-0.487	0.808	0.747	-0.77

Table 10: Average covariates at the contest level: Randomization results

Notes: An observation is a contest. 'Invited players' is the pool of players who were invited to enter a competition, and 'Entrants' is the pool of players who submitted at least one submission during the competition. Treated contests are the contests where a leaderboard was displayed. All variables are defined at the contest level as follows: 'participated_past' is the share of players who have participated in a prediction contest in the past, 'software_code' is the share of players who know how to use a statistical software, 'stat_tools' is the share of players who have statistical skills, 'mach_learning' is the share of players who have machine learning skills, and 'regression' is the share of players who have regression analysis skills.

about participants such as past experience with online competitions and data analysis. There were also asked to create a Kaggle username. With this pool of potential players, we formed 44 competitions of 5 players each. Participants were randomly allocated to these 44 competitions. On average, 3.227 and 3.545 players submitted at least one submission during the competition in the control-group and treatment-group contests, respectively, and this difference is not statistically significant (t-stat= 1.151).

Table 10 shows the outcome of the randomization. The left panel ("Invited players") shows the balance of covariates across competitions in the treatment and control groups. The table shows no statistically significant differences across groups in a number of covariates related to the participants' knowledge of statistical tools and experience. The right panel ("Entrants") repeats the analysis, but restricts attention to the participants who submitted at least one solution during the competition. Again, we find no statistically significant differences across contests in the composition of the control and treatment groups.

	(1)	(2)	(3)	(4)	(5)	(6)
	Maximum score			Number of submissions		
No leaderboard	-0.057**	-0.045**	-0.050**	-31.636***	-26.767***	-26.758***
	(0.022)	(0.019)	(0.020)	(7.073)	(5.250)	(5.973)
	[0.014]	[0.03]	[0.042]	[0.000]	[0.000]	[0.000]
Entrants		0.038***	0.038***		15.302***	15.843***
		(0.013)	(0.014)		(3.050)	(3.244)
Controls	No	No	Yes	No	No	Yes
Observations	44	44	44	44	44	44
R^2	0.135	0.332	0.398	0.323	0.565	0.610
Dep. variable mean	0.192	0.192	0.192	23.636	23.636	23.636

Table 11: The effect of the leaderboard on contest outcomes: Experimental results

Notes: Robust standard errors in parentheses. p-values for Monte Carlo permutation tests to allow for arbitrary randomization procedures in squared brackets (based on 1,000 replications). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a contest. The definition of the variables is as follows: 'No leaderboard' is an indicator for contests without a leaderboard and 'Entrants' is the number of entrants. Controls include the share of participants in the contest who have i) participated in a prediction contest in the past, ii) know how to use a statistical software, iii) have statistical skills, iv) have machine learning skills, and v) have regression analysis skills.

6.2 Experimental Results

Table 11 shows the main results of the experiment, which are in line our model-based evidence: Participation and the maximum score improve when a real-time leaderboard is displayed. Columns 1, 2 and 3 in Table 11 show that outcomes on average worsen in competitions that do not display a leaderboard, relative to competitions with a leaderboard. Column 1 shows that the maximum score was on average 0.057 lower in competitions without a leaderboard, a magnitude that is 29.68 percent of the average maximum score across all contests. This result is robust to controlling for the number of entrants in each competition (column 2) and after controlling for player covariates (column 3).

Columns 4, 5, and 6 in Table 11 show that the number of submissions is on average lower in competitions without a leaderboard versus competitions with a leaderboard. Column 4 shows that competitions without a leaderboard received an average of 31.636

fewer submissions than competitions with a leaderboard, which is a large effect relative to the average number of submissions across all contests. This result is also robust to controlling for the number of entrants in each competition (column 5) and player covariates (column 6).

7 Discussion

We contribute to the literature of dynamic competition design by investigating whether outcomes in a competition improve when players' performance is disclosed in a real-time public leaderboard. We first use field data from Kaggle.com to build and estimate a structural model for the observed design, which features a real-time public leaderboard. We use these parameters to compute the counterfactual equilibrium in which players do not observe a leadeboard.

The comparison of both equilibria shows that a public leadearboard on average improves outcomes: the total number of submissions on average increases 21 percent and the maximum score on average increases 1.7 percent in competitions that display a leaderboard. Even more, we find that a leaderboard on average increases participation because it encourages high-type players to stay longer in the competition, and prompts low-type players to quit earlier.

Displaying a leaderboard, however, may be detrimental in some competitions. A leaderboard may improve or it may worsen outcomes depending on features of a competition such as the cost-to-prize ratio and the shape of the distribution of scores. The reason is that disclosing information discourages players that lag behind. This is, a player may decide to quit when her scores are too low compared to the scores of the players leading the competition. In contrast, without a leaderboard a player's own scores and her beliefs about final scores by her rivals are used to decide when to quit. A leaderboard has theoretically ambiguous effects. We find that the cost-to-prize ratio and the variance of the distribution of scores of high-type players are positively correlated with the difference in the number of submissions with and without a leaderboard. Intuitively, information is more valuable for players when new submissions are more costly, and players are less discouraged from information disclosure when the variance of scores is higher, because the leader of the competition is replaced more frequently. This suggests that a contest designer would benefit from a displaying public leaderboard in competitions where there is a large variation in players' scores or where new submissions are costly relative to the prize.

To complement our model-based analysis, we ran an experiment in which we randomly allocated participants into competitions without a leaderboard (control group) or with a leaderboard (treatment group). The experimental analysis is independent of our modeling assumptions, so it serves the double purpose of providing experimental evidence and testing the predictions of our model. Our experimental findings are consistent with our model-based analysis results: the number of submissions and the maximum score on average increase in competitions when displaying a public leaderboard.

8 References

Aguirregabiria, Victor and Pedro Mira (2007) "Sequential estimation of dynamic discrete games," *Econometrica*, Vol. 75, pp. 1–53.

Aoyagi, Masaki (2010) "Information feedback in a dynamic tournament," *Games and Economic Behavior*, Vol. 70, pp. 242–260.

Arcidiacono, Peter, Patrick Bayer, Jason R Blevins, and Paul B Ellickson (2016) "Estimation of dynamic discrete choice models in continuous time with an application to retail competition," *The Review of Economic Studies*, Vol. 83, pp. 889–931.

Athanasopoulos, George and Rob J Hyndman (2011) "The value of feedback in forecasting competitions," *International Journal of Forecasting*, Vol. 27, pp. 845–849.

Azmat, Ghazala and Marc Möller (2009) "Competition among contests," *The RAND Journal of Economics*, Vol. 40, pp. 743–768.

Bajari, Patrick and Ali Hortacsu (2003) "The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions," *RAND Journal of Economics*, pp. 329–355.

Balafoutas, Loukas, E Glenn Dutcher, Florian Lindner, and Dmitry Ryvkin (2017) "The Optimal Allocation of Prizes in Tournaments of Heterogeneous Agents," *Economic Inquiry*, Vol. 55, pp. 461–478.

Barkley, Aaron, Joachim R Groeger, and Robert A Miller (2019) "Bidding Frictions in Ascending Auctions."

Benkert, Jean-Michel and Igor Letina (2016) "Designing dynamic research contests," *University of Zurich, Department of Economics, Working Paper*.

Bhattacharya, Vivek (2016) "An Empirical Model of R&D Procurement Contests: An Analysis of the DOD SBIR Program," *MIT, Department of Economics, Working Paper*.

Bimpikis, Kostas, Shayan Ehsani, and Mohamed Mostagir (2014) "Designing dynamic contests," *Working paper, Stanford University*.

Boudreau, Kevin J, Nicola Lacetera, and Karim R Lakhani (2011) "Incentives and problem uncertainty in innovation contests: An empirical analysis," *Management Science*, Vol. 57, pp. 843–863.

Boudreau, Kevin J, Karim R Lakhani, and Michael Menietti (2016) "Performance responses to competition across skill levels in rank-order tournaments: field evidence and implications for tournament design," *The RAND Journal of Economics*, Vol. 47, pp. 140–165.

Che, Yeon-Koo and Ian Gale (2003) "Optimal design of research contests," *The American Economic Review*, Vol. 93, pp. 646–671.

Chesbrough, Henry, Wim Vanhaverbeke, and Joel West (2006) *Open innovation: Researching a new paradigm*: Oxford University Press on Demand.

Clark, Derek J and Tore Nilssen (2013) "Learning by doing in contests," *Public Choice*, Vol. 156, pp. 329–343.

Coey, Dominic, Bradley J Larsen, and Brennan C Platt (2019) "Discounts and Deadlines in Consumer Search."

Cohen, Chen, Todd R Kaplan, and Aner Sela (2008) "Optimal rewards in contests," *The RAND Journal of Economics*, Vol. 39, pp. 434–451.

Ding, Wei and Elmar G Wolfstetter (2011) "Prizes and lemons: procurement of innovation under imperfect commitment," *The RAND Journal of Economics*, Vol. 42, pp. 664–680.

Ederer, Florian (2010) "Feedback and motivation in dynamic tournaments," *Journal of Economics & Management Strategy*, Vol. 19, pp. 733–769.

Fullerton, Richard L and R Preston McAfee (1999) "Auctioning entry into tournaments," *Journal of Political Economy*, Vol. 107, pp. 573–605.

Genakos, Christos and Mario Pagliero (2012) "Interim rank, risk taking, and performance in dynamic tournaments," *Journal of Political Economy*, Vol. 120, pp. 782–813.

Goltsman, Maria and Arijit Mukherjee (2011) "Interim performance feedback in multistage tournaments: The optimality of partial disclosure," *Journal of Labor Economics*, Vol. 29, pp. 229–265.

Gross, Daniel P (2015) "Creativity Under Fire: The Effects of Competition on Creative Production," *Available at SSRN 2520123.*

——— (2017) "Performance feedback in competitive product development," *The RAND Journal of Economics*, Vol. 48, pp. 438–466.

Halac, Marina, Navin Kartik, and Qingmin Liu (2014) "Contests for experimentation," *Journal of Political Economy, Forthcoming.*

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) "The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics."

Hendricks, Kenneth and Robert H Porter (1988) "An empirical study of an auction with asymmetric information," *The American Economic Review*, pp. 865–883.

Hinnosaar, Toomas (2017) "Dynamic common-value contests."

Hotz, V Joseph and Robert A Miller (1993) "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, Vol. 60, pp. 497–529.

Hotz, V Joseph, Robert A Miller, Seth Sanders, and Jeffrey Smith (1994) "A simulation estimator for dynamic models of discrete choice," *The Review of Economic Studies*, Vol. 61, pp. 265–289.

Huang, Yan, Param Vir Singh, and Kannan Srinivasan (2014) "Crowdsourcing new product ideas under consumer learning," *Management science*, Vol. 60, pp. 2138–2159.

Jeppesen, Lars Bo and Karim R Lakhani (2010) "Marginality and problem-solving effectiveness in broadcast search," *Organization science*, Vol. 21, pp. 1016–1033.

Kireyev, Pavel (2016) "Markets for Ideas: Prize Structure, Entry Limits, and the Design of Ideation Contests," *HBS, Working Paper*.

Klein, Arnd Heinrich and Armin Schmutzler (2016) "Optimal effort incentives in dynamic tournaments," *Games and Economic Behavior*.

Krasnokutskaya, Elena and Katja Seim (2011) "Bid preference programs and participation in highway procurement auctions," *The American Economic Review*, Vol. 101, pp. 2653–2686.

Lakhani, Karim R, Kevin J Boudreau, Po-Ru Loh, Lars Backstrom, Carliss Baldwin, Eric Lonstein, Mike Lydon, Alan MacCormack, Ramy A Arnaout, and Eva C Guinan (2013) "Prize-based contests can provide solutions to computational biology problems," *Nature biotechnology*, Vol. 31, pp. 108–111.

Landers, Richard N, Kristina N Bauer, and Rachel C Callan (2017) "Gamification of task performance with leaderboards: A goal setting experiment," *Computers in Human Behavior*, Vol. 71, pp. 508–515.

Landers, Richard N and Amy K Landers (2014) "An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance," *Simulation & Gaming*, Vol. 45, pp. 769–785.

Lazear, Edward P and Sherwin Rosen (1979) "Rank-order tournaments as optimum labor contracts."

Lerner, Josh and Jean Tirole (2002) "Some simple economics of open source," *The journal of industrial economics*, Vol. 50, pp. 197–234.

Levin, Dan and James L Smith (1994) "Equilibrium in auctions with entry," *The American Economic Review*, pp. 585–599.

Li, Tong, Isabelle Perrigne, and Quang Vuong (2002) "Structural estimation of the affiliated private value auction model," *RAND Journal of Economics*, pp. 171–193.

Manski, Charles F and Steven R Lerman (1977) "The estimation of choice probabilities from choice based samples," *Econometrica: Journal of the Econometric Society*, pp. 1977–1988.

Megidish, Reut and Aner Sela (2013) "Allocation of Prizes in Contests with Participation Constraints," *Journal of Economics & Management Strategy*, Vol. 22, pp. 713–727.

Moldovanu, Benny and Aner Sela (2001) "The optimal allocation of prizes in contests," *American Economic Review*, pp. 542–558.

——— (2006) "Contest architecture," *Journal of Economic Theory*, Vol. 126, pp. 70–96.

Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007) "Contests for status," *Journal of Political Economy*, Vol. 115, pp. 338–363.

Olszewski, Wojciech and Ron Siegel (2015) "Effort-Maximizing Contests."

Rieck, Thomas (2010) "Information disclosure in innovation contests,"Technical report, Bonn Econ Discussion Papers.

Sisak, Dana (2009) "Multiple-prize contests–the optimal allocation of prizes," *Journal of Economic Surveys*, Vol. 23, pp. 82–114.

Strack, Philipp (2016) "Risk-Taking in Contests: The Impact of Fund-Manager Compensation on Investor Welfare."

Takahashi, Yuya (2015) "Estimating a war of attrition: The case of the U.S. movie theater industry," *The American Economic Review*, Vol. 105, pp. 2204–2241.

Taylor, Curtis R (1995) "Digging for golden carrots: an analysis of research tournaments," *The American Economic Review*, pp. 872–890.

Terwiesch, Christian and Yi Xu (2008) "Innovation contests, open innovation, and multiagent problem solving," *Management science*, Vol. 54, pp. 1529–1543.

Xiao, Jun (2016) "Asymmetric all-pay contests with heterogeneous prizes," *Journal of Economic Theory*, Vol. 163, pp. 178–221.

Zivin, Joshua S Graff and Elizabeth Lyons (2018) "Can Innovators be Created? Experimental Evidence from an Innovation Contest,"Technical report, National Bureau of Economic Research.

Online Appendix: Not For Publication

Dynamic Tournament Design: Evidence from Prediction Contests

Jorge Lemus and Guillermo Marshall

A Additional Tables and Figures

Public Ranking of Winner	Frequency	Probability	Cumulative Probability
1	29	50.88	50.88
2	13	22.81	73.68
3	3	5.26	78.95
4	6	8.77	87.72
5	1	1.75	89.47
6	2	3.51	92.98
11	3	5.26	98.25
54	1	1.75	100.00

Table A.2: Public Leaderboard Ranking of Competition Winners

Note: An observation is a contest.

Number of Competitions	Overall Frequency	Overall Probability	Competitive Frequency	Competitive Probability
1	22,034	71.26	3,556	57.78
2	4,350	14.08	1,024	16.64
3	1,835	5.70	510	8.29
4	908	2.82	275	4.47
5 or more	1,976	6.14	789	12.82

Table A.3: Number of Competitions by User

Note: An observation is a team member.

Contest Number	Name of the Competition	Total Reward	Number of Submissions	Teams	Start Date	Deadline
1	Predict Grant Applications	5,000	2,800	204	12/13/2010	02/20/2011
2	RTA Freeway Travel Time Prediction	10,000	2,958	348	11/23/2010	02/13/2011
3	Deloitte/FIDE Chess Rating Challenge	10,000	1,428	167	02/07/2011	05/04/2011
4	Heritage Health Prize	500,000	23,421	1,221	04/04/2011	04/04/2013
5	Wikipedia's Participation Challenge	10,000	995	88	06/28/2011	09/20/2011
6	Allstate Claim Prediction Challenge	10,000	1,278	102	07/13/2011	10/12/2011
7	dunnhumby's Shopper Challenge	10,000	1,872	277	07/29/2011	09/30/2011
8	Give Me Some Credit	5,000	7,658	920	09/19/2011	12/15/2011
9	Don't Get Kicked!	10,000	7,167	570	09/30/2011	01/05/2012
10	Algorithmic Trading Challenge	10,000	1,169	95	11/11/2011	01/08/2012
11	What Do You Know?	5,000	1,616	228	11/18/2011	02/29/2012
12	Photo Quality Prediction	5,000	1,315	194	10/29/2011	11/20/2011
13	Benchmark Bond Trade Price Challenge	17,500	2,487	248	01/27/2012	04/30/2012
14	KDD Cup 2012, Track 1	8,000	13,076	657	02/20/2012	06/01/2012
15	KDD Cup 2012, Track 2	8,000	5,276	163	02/20/2012	06/01/2012
16	Predicting a Biological Response	20,000	7,668	647	03/16/2012	06/15/2012
17	Online Product Sales	22,500	3,532	346	05/04/2012	07/03/2012
18	EMI Music Data Science Hackathon - July 21st - 24 hours	10,000	1,282	132	07/21/2012	07/22/2012
19	Belkin Energy Disaggregation Competition	25,000	1,399	160	07/02/2013	10/30/2013
20	Merck Molecular Activity Challenge	40,000	2,979	236	08/16/2012	10/16/2012
21	U.S. Census Return Rate Challenge	25,000	2,385	231	08/31/2012	11/11/2012
22	Amazon.com - Employee Access Challenge	5,000	16,872	1,687	05/29/2013	07/31/2013
23	The Marinexplore and Cornell University Whale Detection Challenge	10,000	3,282	244	02/08/2013	04/08/2013
24	See Click Predict Fix - Hackathon	1,000	1,001	79	09/28/2013	09/29/2013
25	KDD Cup 2013 - Author Disambiguation Challenge (Track 2)	7,500	2,216	235	04/19/2013	06/12/2013
26	Influencers in Social Networks	2,350	2,004	129	04/13/2013	04/14/2013
27	Personalize Expedia Hotel Searches - ICDM 2013	25,000	3,409	331	09/03/2013	11/04/2013
28	StumbleUpon Evergreen Classification Challenge	5,000	7,123	593	08/16/2013	10/31/2013
29	Personalized Web Search Challenge	9,000	3,021	177	10/11/2013	01/10/2014
30	See Click Predict Fix	4,000	5,314	517	09/29/2013	11/27/2013
31	Allstate Purchase Prediction Challenge	50,000	24,526	1,568	02/18/2014	05/19/2014
32	Higgs Boson Machine Learning Challenge	13,000	35,772	1,785	05/12/2014	09/15/2014
33	Acquire Valued Shoppers Challenge	30,000	25,138	952	04/10/2014	07/14/2014
34	The Hunt for Prohibited Content	25,000	4,992	285	06/24/2014	08/31/2014
35	Liberty Mutual Group - Fire Peril Loss Cost	25,000	14,751	634	07/08/2014	09/02/2014
36	Tradeshift Text Classification	5,000	4,632	296	10/02/2014	11/10/2014
37	Driver Telematics Analysis	30,000	36,065	1,528	12/15/2014	03/16/2015
38	Diabetic Retinopathy Detection	100,000	7,002	661	02/17/2015	07/27/2015
39	Click-Through Rate Prediction	15,000	27,202	1,417	11/18/2014	02/09/2015
40	Otto Group Product Classification Challenge	10,000	34,300	2,734	03/17/2015	05/18/2015
41	Crowdflower Search Results Relevance	20,000	23,237	1,326	05/11/2015	07/06/2015
42	Avito Context Ad Clicks	20,000	5,317	360	06/02/2015	07/28/2015
43	ICDM 2015: Drawbridge Cross-Device Connections	10,000	2,355	340	06/01/2015	08/24/2015
44	Caterpillar Tube Pricing	30,000	23,834	1,187	06/29/2015	08/31/2015
45	Liberty Mutual Group: Property Inspection Prediction	25,000	40,594	2,054	07/06/2015	08/28/2015
46	Coupon Purchase Prediction	50,000	18,477	1,076	07/16/2015	09/30/2015
47	Springleaf Marketing Response	100,000	34,861	1,914	08/14/2015	10/19/2015
48	Truly Native?	10,000	3,222	274	08/06/2015	10/14/2015
49	Rossmann Store Sales	35,000	58,915	2,861	09/30/2015	12/14/2015
50	Homesite Quote Conversion	20,000	28,571	1,334	11/09/2015	02/08/2016
51	Prudential Life Insurance Assessment	30,000	42,336	2,452	11/23/2015	02/15/2016
52	BNP Paribas Cardif Claims Management	30,000	48,442	2,702	02/03/2016	04/18/2016
53	Home Depot Product Search Relevance	40,000	32,937	1,935	01/18/2016	04/25/2016
54	Santander Customer Satisfaction	60,000	93,031	5,117	03/02/2016	05/02/2016
55	Expedia Hotel Recommendations	25,000	22,709	1,974	04/15/2016	06/10/2016
56	Avito Duplicate Ads Detection	20,000	8,134	548	05/06/2016	07/11/2016
57	Draper Satellite Image Chronology	75,000	2,734	401	04/29/2016	06/27/2016

Table A.1: Summary of the Competitions in the Data (Full List)

Note: The table only considers submissions that received a score. The total reward is measured in US dollars at the moment of the competition.

	μ	SE	λ	SE	σ	SE	$\log L(\hat{\delta})/N$	N
unimelb	2.2669	0.2902	57.5123	1.542	0.016	0.0005	-2.2235	1391
RTA	1.5828	0.1624	55.2645	1.426	0.0135	0.0005	-1.9971	1502
ChessRatings2	2.9122	0.6512	49.9269	2.1465	0.0044	0.0003	-2.4638	541
hhp	2.585	0.1701	191.9182	1.6088	0.0089	0.0004	-3.809	14231
wikichallenge	2.4171	0.5405	63.4633	2.6914	0.0098	0.0005	-2.6282	556
ClaimPredictionChallenge	2.1404	0.5191	70.954	2.9311	0.0018	0.0003	-2.7177	586
dunnhumbychallenge	2.1786	0.3112	43.0186	1.4799	0.0133	0.0005	-1.5467	845
GiveMeSomeCredit	1.8283	0.1175	60.4017	0.9488	0.014	0.0004	-1.9153	4053
DontGetKicked	1.9281	0.1919	87.912	1.5388	0.0064	0.0003	-2.856	3264
AlgorithmicTradingChallenge	4.2848	1.1063	60.5909	2.5423	0.0021	0.0002	-3.0526	568
WhatDoYouKnow	2.5017	0.4906	59.9231	2.2649	0.0069	0.0004	-3.1705	700
PhotoQualityPrediction	2.292	0.367	26.7708	1.1353	0.0044	0.0003	-1.5272	556
benchmark-bond-trade-price-challenge	3.104	0.5173	60.6621	1.7842	0.0103	0.0005	-2.2271	1156
kddcup2012-track1	2.9802	0.2409	99.8329	1.1439	0.0051	0.0003	-3.7538	7617
kddcup2012-track2	2.3884	0.3927	129.954	2.4908	0.0003	0.0001	-3.1547	2722
bioresponse	2.0325	0.1887	79.5422	1.2726	0.0049	0.0003	-2.988	3907
online-sales	2.2939	0.2742	46.2428	1.1645	0.0054	0.0003	-2.3825	1577
MusicHackathon	3.8714	0.5974	18.0368	0.7602	0.0023	0.0002	-1.7897	563
belkin-energy-disaggregation-competition	2.3403	0.6491	118.6425	4.4124	0.0034	0.0003	-3.5879	723
MerckActivity	2.5245	0.3435	52.7299	1.2958	0.0014	0.0002	-2.2825	1656
us-census-challenge	2.1093	0.2723	54.9508	1.5675	0.0226	0.0007	-2.2713	1229
amazon-employee-access-challenge	2.5541	0.1252	49.1849	0.5004	0.0087	0.0003	-2.1492	9663
whale-detection-challenge	2.1498	0.2925	55.8479	1.4282	0.002	0.0002	-2.829	1529
the-seeclickfix-311-challenge	2.5648	0.5884	40.8782	1.9122	0.0015	0.0002	-2.2612	457
kdd-cup-2013-author-disambiguation	2.6885	0.5488	69.1264	2.1666	0.0009	0.0002	-3.1262	1018
predict-who-is-more-influential-in-a-social-network	2.9345	0.476	44.2838	1.4482	0.011	0.0005	-2.0247	935
expedia-personalized-sort	2.3917	0.3088	45.2652	1.1742	0.0033	0.0003	-2.1829	1486
stumbleupon	2.6926	0.209	51.8293	0.7768	0.0059	0.0003	-2.3562	4452
yandex-personalized-web-search-challenge	1.7677	0.2831	116.9086	2.8313	0.0028	0.0003	-3.4408	1705
see-click-predict-fix	1.8173	0.1905	70.48	1.3438	0.0019	0.0002	-2.9384	2751
allstate-purchase-prediction-challenge	1.9856	0.1276	125.4499	1.1689	0.0009	0.0001	-2.6272	11519
higgs-boson	2.3698	0.114	122.1003	0.8177	0.0043	0.0002	-3.6653	22298
acquire-valued-shoppers-challenge	2.0772	0.1316	165.2723	1.2866	0.0032	0.0002	-3.846	16500
avito-prohibited-content	2.5922	0.3953	106.7729	2.0667	0.0009	0.0001	-3.6088	2669
liberty-mutual-fire-peril	3.3432	0.3202	122.5353	1.3388	0.001	0.0001	-3.7003	8377
tradeshift-text-classification	3.1365	0.3983	62.5714	1.2269	0.0007	0.0001	-2.5175	2601
axa-driver-telematics-analysis	2.4434	0.1408	127.4859	0.8925	0.0003	0.0001	-3.3961	20405
diabetic-retinopathy-detection	1.8353	0.2148	102.5382	1.9395	0.0067	0.0004	-2.3459	2795
avazu-ctr-prediction	3.2065	0.1981	109.0486	0.8983	0.0026	0.0002	-3.9573	14735
otto-group-product-classification-challenge	3.2732	0.1397	55.4236	0.3993	0.0005	0.0001	-2.2785	19269
crowdflower-search-relevance	2.2697	0.1163	79.1331	0.6272	0.0047	0.0003	-3.193	15919
avito-context-ad-clicks	1.8973	0.2191	81.359	1.5907	0.0035	0.0003	-3.3652	2616
icdm-2015-drawbridge-cross-device-connections	2.1009	0.3454	57.1866	1.7615	0	0.0002	-2.6684	1054
caterpillar-tube-pricing	3.2701	0.1758	68.2329	0.5562	0.0025	0.0002	-3.0241	15047
liberty-mutual-group-property-inspection-prediction	3.1055	0.1152	67.0227	0.4112	0.0047	0.0002	-2.8525	26573
coupon-purchase-prediction	2.0586	0.1093	73.1853	0.6539	0.005	0.0003	-1.5545	12526
springleaf-marketing-response	3.0405	0.1689	97.3279	0.7153	0.0014	0.0001	-2.9404	18513
dato-native	4.5328	0.7891	50.8066	1.4157	0	0.0001	-2.6522	1288
rossmann-store-sales	2.8735	0.0926	89.6868	0.4478	0.0068	0.0003	-2.5756	40105
homesite-quote-conversion	2.4958	0.1548	128.5381	0.9678	0.0035	0.0002	-3.3492	17638
prudential-life-insurance-assessment	2.16	0.0799	78.0741	0.4707	0.0068	0.0003	-2.7802	27512
bnp-paribas-cardif-claims-management	2.6826	0.0903	63.8265	0.3564	0.0036	0.0002	-3.1061	32069
home-depot-product-search-relevance	2.4158	0.1501	124.638	0.9775	0.0032	0.0002	-3.5879	16258
santander-customer-satisfaction	2.3098	0.0563	75.1579	0.3048	0.0037	0.0002	-3.108	60816
expedia-hotel-recommendations	2.2792	0.086	43.2208	0.3422	0.0042	0.0003	-1.7432	15948
avito-duplicate-ads-detection	3.6084	0.4866	108.5597	1.6678	0.0004	0.0001	-3.0469	4237
draper-satellite-image-chronology	3.0412	0.4177	53.1792	1.4238	0.007	0.0004	-2.6678	1395

Table A.4: Maximum Likelihood Estimates of the Cost and Arrival Distributions

Note: The model is estimated separately for each contest. Asymptotic standard errors are reported in the columns that are labeled 'SE.'

	Type 1			Type 2				
	θ_1^{mean}	$\theta_1^{st.dev}$	κ_1	θ_2^{mean}	$\theta_2^{st.dev}$	κ_2	$\log L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}})/N$	N
unimelb	0.907	0.0469	0.7457	0.3726	0.8017	0.2543	0.8585	1391
RTA	0.6004	0.154	0.679	0.6802	0.4909	0.321	-0.039	1502
ChessRatings2	0.5928	0.0842	0.7497	1.0411	0.1358	0.2503	0.5446	541
hhp	0.6691	0.083	0.657	0.4792	0.5619	0.343	0.3912	14231
wikichallenge	0.7396	0.0976	0.4654	0.617	0.263	0.5346	0.3319	556
ClaimPredictionChallenge	1.4295	0.5305	0.7111	0.3407	0.5064	0.2889	-1.0757	586
dunnhumbychallenge	1.0098	0.2109	0.7482	0.7466	0.2988	0.2518	-0.0454	845
GiveMeSomeCredit	0.519	0.0167	0.7795	0.3599	0.4358	0.2205	2.0494	4053
DontGetKicked	0.7419	0.0699	0.5768	0.7251	0.2224	0.4232	0.6483	3264
AlgorithmicTradingChallenge	0.8133	0.1532	0.7088	0.6172	0.209	0.2912	0.2639	568
WhatDoYouKnow	0.7753	0.1413	0.6434	0.6322	0.253	0.3566	0.2662	700
PhotoQualityPrediction	0.5873	0.0452	0.6808	0.5146	0.1076	0.3192	1.3044	556
benchmark-bond-trade-price-challenge	0.7618	0.1422	0.2868	0.8932	0.3061	0.7132	-0.1002	1156
kddcup2012-track1	0.6891	0.1721	0.7095	0.5523	0.7572	0.2905	-0.2329	7617
kddcup2012-track2	0.8914	0.151	0.4495	0.6644	0.2643	0.5505	0.0459	2722
bioresponse	0.7804	0.1344	0.5065	0.6357	0.2731	0.4935	0.1555	3907
online-sales	0.8445	0.118	0.6402	0.6456	0.2928	0.3598	0.3067	1577
MusicHackathon	0.9258	0.1397	0.7245	0.6437	0.247	0.2755	0.2255	563
belkin-energy-disaggregation-competition	0.4018	0.358	0.5898	1.3458	1.7243	0.4102	-1.0911	723
MerckActivity	0.6978	0.1448	0.5714	0.5709	0.245	0.4286	0.2189	1656
us-census-challenge	0.8083	0.2064	0.1507	0.916	0.5135	0.8493	-0.6594	1229
amazon-employee-access-challenge	0.7551	0.0567	0.6302	0.545	0.4323	0.3698	0.6967	9663
whale-detection-challenge	0.6904	0.0603	0.5873	0.6346	0.1204	0.4127	1.0703	1529
the-seeclickfix-311-challenge	0.954	0.0386	0.2673	0.7401	0.2573	0.7327	0.2535	457
kdd-cup-2013-author-disambiguation	1.5339	0.2196	0.2452	0.8268	0.5049	0.7548	-0.7533	1018
predict-who-is-more-influential-in-a-social-network	0.6743	0.0559	0.5779	0.6192	0.1644	0.4221	0.9089	935
expedia-personalized-sort	0.9385	0.2123	0.6802	0.7981	0.5742	0.3198	-0.2242	1486
stumbleupon	0.6369	0.0678	0.6529	0.5653	0.288	0.3471	0.7431	4452
yandex-personalized-web-search-challenge	1.0392	0.7247	0.9036	0.7315	0.2634	0.0964	-1.0347	1705
see-click-predict-fix	0.7567	0.0543	0.6962	0.5926	0.3231	0.3038	0.9078	2751
allstate-purchase-prediction-challenge	0.5038	0.0104	0.8313	0.4071	0.2553	0.1687	2.6755	11519
higgs-boson	0.6632	0.0738	0.7253	0.4505	0.463	0.2747	0.6744	22298
acquire-valued-shoppers-challenge	0.8254	0.1894	0.4967	0.5916	0.545	0.5033	-0.3439	16500
avito-prohibited-content	0.4834	0.0181	0.6585	0.4445	0.0691	0.3415	2.1097	2669
liberty-mutual-fire-peril	0.7387	0.1754	0.6447	0.5652	0.3024	0.3553	0.0656	8377
tradeshift-text-classification	0.8188	0.141	0.4289	0.6122	0.2731	0.5711	0.0676	2601
axa-driver-telematics-analysis	1.0063	0.1863	0.4779	0.8082	0.4058	0.5221	-0.1938	20405
diabetic-retinopathy-detection	1.2998	0.3702	0.7685	0.783	0.5386	0.2315	-0.6156	2795
avazu-ctr-prediction	0.5778	0.1829	0.803	0.699	0.3436	0.197	0.2071	14735
otto-group-product-classification-challenge	0.7953	0.1462	0.5697	0.6681	0.3331	0.4303	0.0973	19269
crowdflower-search-relevance	0.6626	0.1138	0.5625	0.51	0.3922	0.4375	0.1954	15919
avito-context-ad-clicks	1.0044	0.5639	0.7838	0.7382	0.2659	0.2162	-0.7511	2616
icdm-2015-drawbridge-cross-device-connections	1.4725	0.058	0.166	1.0204	0.3095	0.834	-0.2142	1054
caterpillar-tube-pricing	0.5999	0.0491	0.6027	0.4656	0.3495	0.3973	0.8037	15047
liberty-mutual-group-property-inspection-prediction	0.7789	0.1128	0.4056	0.5568	0.362	0.5944	0.2598	26573
coupon-purchase-prediction	0.7612	0.0031	0.1434	0.6923	0.5603	0.8566	0.2596	12526
springleaf-marketing-response	0.8079	0.1112	0.5603	0.6637	0.3488	0.4397	0.2251	18513
dato-native	0.6084	0.0572	0.7182	0.6883	0.0128	0.2818	1.5161	1288
rossmann-store-sales	0.65	0.1158	0.5893	0.4968	0.3264	0.4107	0.3807	40105
homesite-quote-conversion	0.5534	0.0647	0.5543	0.5752	0.2611	0.4457	0.6504	17638
prudential-life-insurance-assessment	0.7326	0.066	0.6366	0.5313	0.3874	0.3634	0.6549	27512
bnp-paribas-cardif-claims-management	0.5812	0.1488	0.638	0.5601	0.488	0.362	-0.0302	32069
home-depot-product-search-relevance	0.6521	0.2161	0.5194	0.8518	0.4976	0.4806	-0.4202	16258
santander-customer-satisfaction	0.4604	0.0949	0.785	0.326	0.3797	0.215	2.2061	60816
expedia-hotel-recommendations	0.4196	0.5705	0.3425	0.6816	0.0123	0.6575	1.8104	15948
avito-duplicate-ads-detection	0.928	0.0937	0.6124	0.7649	0.329	0.3876	0.4461	4237
draper-satellite-image-chronology	0.0118	0.3054	0.4991	1.4943	1.0653	0.5009	-1.24	1395

Table A.5: EM Algorithm Estimates for the Type-specific Distribution of Scores, q_θ.

Note: The model is estimated separately for each contest. $\theta_i^{st.dev}$ and $\theta_i^{st.dev}$ are the parameters in type i's distribution of scores $Q_i(s) = \Phi\left(\frac{s-\theta_i^{st.dev}}{\theta_i^{st.dev}}\right)$. κ_i is the fraction of players of type θ_i. $\log L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}})/N$ is the value of the log-likelihood function evaluated at the EM estimates. Standard errors are available.

	α	SE	β	SE	$\log L(\hat{\delta})/N$
unimelb	-0.0055	0.0305	0.993	0.0316	1391
RTA	-0.0479	0.0298	0.9651	0.0363	1502
ChessRatings2	-0.0121	0.0799	1.0182	0.1064	541
hhp	-0.0055	0.009	0.9941	0.0102	14231
wikichallenge	-0.0052	0.0477	1.0017	0.0593	556
ClaimPredictionChallenge	0.0477	0.0512	0.8792	0.0416	586
dunnhumbychallenge	0.003	0.046	0.9766	0.0471	845
GiveMeSomeCredit	0.0014	0.017	0.9961	0.0224	4053
DontGetKicked	0.0065	0.0233	0.9979	0.0285	3264
AlgorithmicTradingChallenge	0.0801	0.0511	0.7757	0.0595	568
WhatDoYouKnow	-0.0062	0.0543	1.0049	0.0686	700
PhotoQualityPrediction	0.0041	0.0547	0.9981	0.0807	556
benchmark-bond-trade-price-challenge	-0.0207	0.0404	1.0488	0.0464	1156
kddcup2012-track1	-0.2673	0.0119	0.7731	0.0122	7617
kddcup2012-track2	0.0045	0.0212	1.0108	0.0243	2722
bioresponse	-0.0324	0.0188	0.9517	0.0221	3907
online-sales	0.0049	0.0316	0.9882	0.0366	1577
MusicHackathon	-0.0002	0.0514	1.0002	0.0551	563
belkin-energy-disaggregation-competition	-0.29	0.038	0.4271	0.02	723
MerckActivity	0.0049	0.0265	0.9924	0.0305	1656
us-census-challenge	0.0001	0.0324	0.998	0.0326	1229
amazon-employee-access-challenge	-0.0084	0.0121	1.0035	0.0148	9663
whale-detection-challenge	-0.0009	0.033	1.002	0.044	1529
the-seeclickfix-311-challenge	-0.0027	0.0598	1.002	0.0689	457
kdd-cup-2013-author-disambiguation	-0.0002	0.0369	1.0014	0.0321	1018
predict-who-is-more-influential-in-a-social-network	-0.0185	0.0404	1.0231	0.0568	935
expedia-personalized-sort	0.0004	0.0333	1.0024	0.0341	1486
stumbleupon	-0.0134	0.0169	0.9552	0.0222	4452
yandex-personalized-web-search-challenge	-0.0083	0.0258	1.0035	0.0238	1705
see-click-predict-fix	-0.0004	0.0217	1.0004	0.0255	2751
allstate-purchase-prediction-challenge	-0.002	0.0105	0.9994	0.015	11519
higgs-boson	-0.002	0.0074	0.9935	0.0089	22298
acquire-valued-shoppers-challenge	-0.0058	0.0082	0.9962	0.0087	16500
avito-prohibited-content	-0.0002	0.0215	1	0.0306	2669
liberty-mutual-fire-peril	-0.0057	0.0121	0.9613	0.0144	8377
tradeshift-text-classification	0.0002	0.022	0.9969	0.026	2601
axa-driver-telematics-analysis	-0.0013	0.0082	1.0003	0.0085	20405
diabetic-retinopathy-detection	-0.0017	0.0296	0.9984	0.0248	2795
avazu-ctr-prediction	0.0006	0.0096	0.9993	0.0127	14735
otto-group-product-classification-challenge	0.0024	0.0084	0.9968	0.01	19269
crowdflower-search-relevance	-0.0005	0.0085	0.9885	0.0104	15919
avito-context-ad-clicks	0.0054	0.0217	0.9988	0.0196	2616
icdm-2015-drawbridge-cross-device-connections	-0.0007	0.0436	1.0023	0.0413	1054
caterpillar-tube-pricing	0.0012	0.0086	0.9997	0.0104	15047
liberty-mutual-group-property-inspection-prediction	-0.0108	0.0066	0.9862	0.0078	26573
coupon-purchase-prediction	-0.0187	0.0094	0.9515	0.0092	12526
springleaf-marketing-response	0.0042	0.0086	1.0064	0.0101	18513
dato-native	-0.0088	0.0404	1.0015	0.0592	1288
rossmann-store-sales	-0.0475	0.0055	0.9825	0.0071	40105
homesite-quote-conversion	-0.0005	0.0084	0.9932	0.011	17638
prudential-life-insurance-assessment	-0.0054	0.0071	0.9949	0.0094	27512
bnp-paribas-cardif-claims-management	-0.0011	0.006	0.9959	0.0072	32069
home-depot-product-search-relevance	0.0013	0.0092	1.0022	0.0096	16258
santander-customer-satisfaction	0.0008	0.0044	1	0.0062	60816
expedia-hotel-recommendations	-0.0006	0.0084	0.9996	0.0102	15948
avito-duplicate-ads-detection	-0.0017	0.0191	0.999	0.0216	4237
draper-satellite-image-chronology	-0.1409	0.0283	0.6141	0.0202	1395

Table A.6: Maximum Likelihood Estimates of the Distribution of Private Scores Conditional on Public Scores.

Note: The Conditional Distribution is Assumed to be Given by $p^{private} = \alpha + \beta p^{public} + \epsilon$, with ϵ Distributed According to a Double Exponential Distribution. The model is estimated separately for each contest. Asymptotic standard errors are reported in the columns that are labeled 'SE.'

	(1)	(2)
	Total effect $(n_{lb} - n_{no\ lb})$	Information Disclosure $(n_{lb} - n'_{no\ lb})$
σ	3.1115**	4.1751***
	(1.5024)	(1.5132)
μ	-1.1838	0.1040
	(1.9585)	(1.9743)
λ	1.5729	1.1141
	(1.6473)	(1.6273)
$\kappa_{high\ type}$	-0.5753	-1.2530
	(2.7958)	(3.0412)
$\theta^{st.dev}_{high\ type}$	4.9351**	5.1694**
	(2.0355)	(2.0304)
$\theta^{st.dev}_{low\ type}$	0.3288	0.4106
	(2.8737)	(2.9349)
$\theta^{mean}_{high\ type}$	-3.5381	-3.4160
	(3.8722)	(3.9662)
$\theta^{mean}_{low\ type}$	-3.3526*	-3.2002*
	(1.7289)	(1.7808)
Observations	57	57
R^2	0.285	0.298

Table A.7: Explaining the impact of the leaderboard on contest outcomes

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An observation is a contest. Total effect is the difference in the average number of submissions by a team with a leaderboard and without a leaderboard. Information disclosure is the difference between the average number of submissions by a team in the equilibrium with a leaderboard and the average number of submissions by the same team when this team that does not observe the leaderboard but knows that all other teams observe the leaderboard. All independent variables are measured in standard deviations. σ is the parameter of the cost distribution; μ is the arrival rate of new players; λ is the parameter of the distribution of time between submissions; $\kappa_{high\ type}$ is the share of high-type players; $\theta_j^{st.dev}$ and θ_j^{mean} are the standard deviation and mean of the distribution of scores of players of type θ_j. Type θ_i is defined as the high-type if $\theta_i^{mean} + 3\theta_i^{st.dev} > \theta_j^{mean} + 3\theta_j^{st.dev}$.

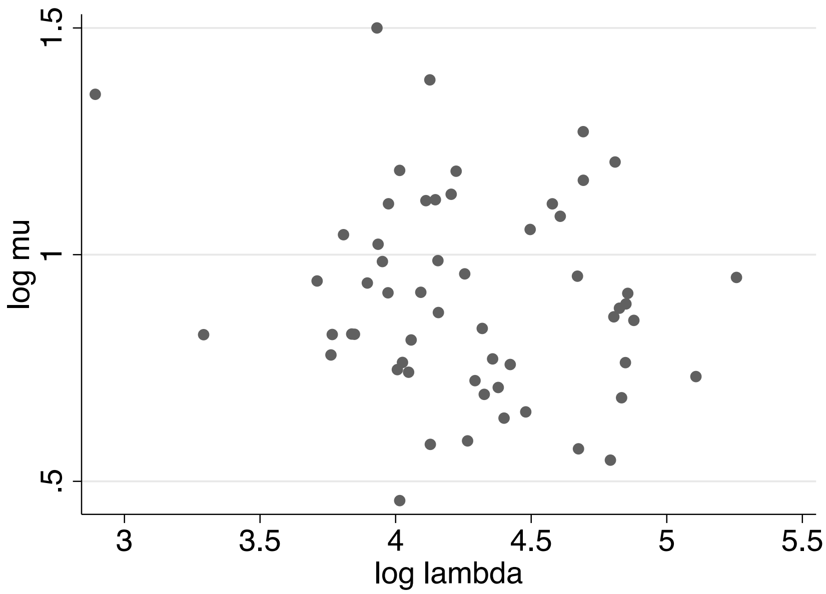


Figure A.1: Correlation between μ (arrival rate of new teams) and λ (arrival rate of new submissions)

Note: The estimates of λ and μ are reported in Table A.4. The coefficient of correlation between λ and μ is -0.18.

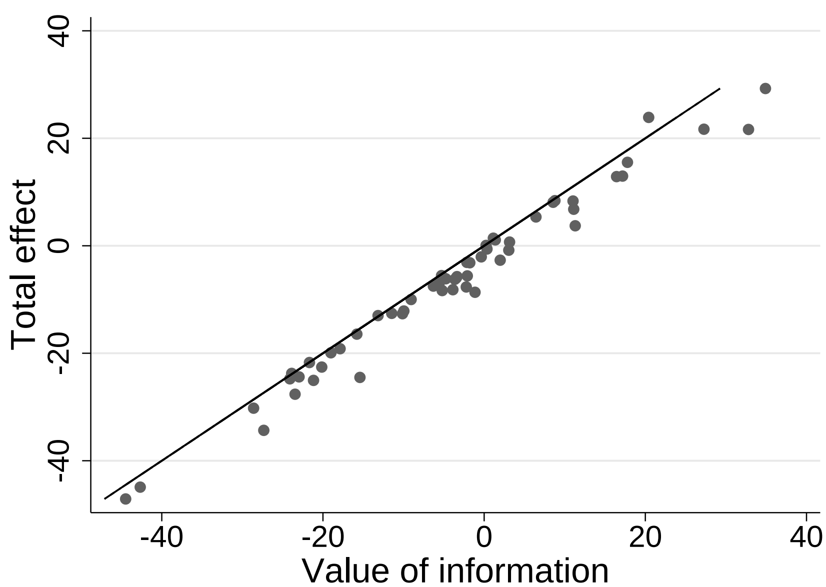


Figure A.2: Value of Information versus total effect of hiding a leaderboard

B Estimation Details

In this section, we provide an overview of the estimation procedure used to estimate the primitives of the model.

B.1 Distribution of entry times

We assume that the time at which a player enters a competition follows an exponential distribution with a contest-specific parameter, μ. Given the vector of entry times for the set of players I in a given contest, $\{t_i\}_{i\in I}$, we estimate μ by using the maximum likelihood estimator:

$$\hat{\mu} = \arg\max \log L(\mu) = \arg\max \sum_{i\in I} \log(\mu) - \mu t_i = \frac{1}{\bar{t}},$$

where $\bar{t} = \sum_{i\in I} t_i/|I|$.

B.2 Distribution of time between submissions

We assume that the time between submissions follows an exponential distribution with a contest-specific parameter, λ. Given the vector of times between submissions for the set of players I in a given contest, $\{t_{i,m}\}_{m\in M_i, i\in I}$, we estimate λ by using the maximum likelihood estimator:

$$\hat{\lambda} = \arg\max \log L(\lambda) = \arg\max \sum_{i\in I} \sum_{m\in M_i} \log(\lambda) - \lambda t_{i,m} = \frac{1}{\bar{t}},$$

where $\bar{t} = \sum_{i\in I} \sum_{m\in M_i} t_{i,m}/|\{t_{i,m}\}_{m\in M_i, i\in I}|$.

B.3 Type-specific distribution of scores

We assume that each player is of one of two types: θ_1 or θ_2. The fraction of players of type θ_j is κ_j, with $\kappa_1+\kappa_2 = 1$. Each type j draws scores from a type-specific distribution of scores, $f(\cdot|\theta_j)$. We assume $f(\cdot|\theta_j)$ is the density function of a normal distribution with mean θ_j^{mean} and standard deviation $\theta_j^{st.dev}$. We define $\boldsymbol{\theta} = \{\theta_j = (\theta_j^{mean}, \theta_j^{st.dev})\}_{j=1,2}$ and $\boldsymbol{\kappa} = \{\kappa_j\}_{j=1,2}$.

Using the distributions $f(\cdot|\theta_1)$ and $f(\cdot|\theta_2)$ and player i's observed scores $s_i = \{s_{i1}, \ldots, s_{iM_i}\}$, the posterior probability that player i is of type $\theta_j \in \Theta$ is given by

$$h(\theta_j|s_i) = \frac{\kappa_j \prod_{m=1}^{M_i} f(s_{im}|\theta_j)}{\sum_{k\in\Theta} \kappa_k \prod_{m=1}^{M_i} f(s_{im}|\theta_k)}, \tag{5}$$

where we make use of Bayes' identity.

We use the EM algorithm to find the maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$. The EM algorithm is an iterative method, where each iteration consists of computing an expectation and then maximizing the expectation with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$, and these steps are repeated until convergence of the vector of parameters (see Hastie et al. 2009 for more details).

Iteration $t+1$ makes use of the estimates of iteration t: $\boldsymbol{\theta}^t, \boldsymbol{\kappa}^t$. The expectation is given by

$$\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{\kappa}|\boldsymbol{\theta}^t, \boldsymbol{\kappa}^t) \;=\; \sum_i \sum_{\theta_k\in\Theta} \sum_{m\in M_i} h(\theta_k|s_i, \boldsymbol{\theta}^t, \boldsymbol{\kappa}^t) \log(\kappa_k f(s_{im}|\theta_k)).$$

Given $(\boldsymbol{\theta}^t, \boldsymbol{\kappa}^t)$, $\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{\kappa}|\boldsymbol{\theta}^t, \boldsymbol{\kappa}^t)$ has a unique maximum given by $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\kappa}^{t+1})$. Given our assumptions, one can obtain an analytic solution for $(\boldsymbol{\theta}^{t+1}, \boldsymbol{\kappa}^{t+1})$. The estimates of the model are obtained by iterating over the expectation and maximization steps until convergence of the estimates: $\rho((\boldsymbol{\theta}^t, \boldsymbol{\kappa}^t), (\boldsymbol{\theta}^{t+1}, \boldsymbol{\kappa}^{t+1})) < \varepsilon$, where $\rho(\cdot)$ is the Euclidean metric. We use a tolerance level of 1E-8.[25]

When implementing the EM algorithm, we only use three scores for each player i in a given contest. We restrict the set of scores s_i of player i to player i's median score, player i's 75th percentile score, and player i's maximum score.

[25] Alternatively, we can iterate until the log-likelihood converges.

B.3.1 Classifying players

For the purposes of our empirical model, we assign a type to each player of each contest in the data. We classify players using the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$, which fully characterize the distribution of types and the distribution of scores of each type, and the vector of scores of each player. Specifically, we use the posterior probability of player i being of type θ_j in Equation 5, which makes use of the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ and the vector of scores of player i, s_i, which we define as player i's median score, player i's 75th percentile score, and player i's maximum score. Using these posterior probabilities, we classify player i as being of type θ_1 if $h(\theta_1|s_i) > h(\theta_2|s_i)$ and of type 2 otherwise. That is, we use the Bayes classifier, which minimizes expected prediction error (Hastie et al., 2009).

B.4 Conditional distribution of the private scores

We assume that the relationship between private and public scores is given by $p^{private} = \alpha + \beta p^{public} + \varepsilon$, where ε is distributed according to a standard double exponential distribution, and α and β are contest-specific parameters. Given the pairs of scores for all M submissions in a contest, $\{(p_m^{public}, p_m^{private})\}_{m\in M}$, we estimate (α, β) by using the maximum likelihood estimator:

$$(\hat{\alpha}, \hat{\beta}) = \arg\max \log L(\alpha, \beta) = \arg\max \sum_{m\in M} -\varepsilon_m + \exp\{-\varepsilon_m\},$$

where $\varepsilon_m = p_m^{private} - \alpha - \beta p_m^{public}$.

B.5 Estimating the distribution of costs

Players decide to build a new submission if the benefits are greater than the cost of building a new submission. The distribution of costs is identified based on how variation in the benefits of building a new submission changes the decision of a player to build a new submission. We compute the benefits of building a new submission via simulation, and use these simulated benefits to compute the maximum likelihood estimates of the cost distribution.

Simulating the benefits of building a new submission require estimates of all primitives of the model except for the distribution of costs as well as estimates of the equilibrium conditional choice probabilities (CCPs). We first discuss how we estimate the CCPs and then discuss how we simulate the benefits of building a new submission.

B.5.1 Conditional Choice Probabilities (CCPs)

The object of interest is $\Pr(\text{play}|s,t,\theta)$, where s is a vector of state variables, t is time, and θ is player type. We estimate these probabilities using data on both decisions to build a new submission and state variables (e.g., scores on the leaderboard). At this point, we make some assumptions about the information in the public leaderboard that players consider relevant. Specifically, we assume that a player at time t keeps track of time, her own maximum score up to time t, and the top 10 highest scores on the leaderboard up to that time. We found that incorporating additional variables added little explanatory power to our model.

We then approximate the CCPs using the function

$$\Pr(\text{play}|s,t,\theta) = \frac{\exp\{f(s,t,\delta_\theta)\}}{1+\exp\{f(s,t,\delta_\theta)\}}, \tag{6}$$

where $f(\cdot)$ is a flexible function of s and t, and θ, and δ_θ is a vector of parameters that is specific to player type θ. We estimate $\{\delta_\theta\}_{\theta\in\Theta}$ using a maximum likelihood estimator. We model $f(s,t,\delta_\theta)$ as

$$\begin{aligned} f(s,t,\delta_\theta) &= [1, t, t^2, t^3, \mathit{team\_max\_score}, \mathit{team\_max\_score}^2, \mathit{team\_max\_score}\cdot t, \\ &\quad Y, Y*Y, Y\cdot t, \mathit{top}1, \mathit{top}3, \mathit{top}10, \mathit{top}1\cdot t, \mathit{top}3\cdot t, \mathit{top}10\cdot t]'\delta_\theta, \end{aligned}$$

where $\mathit{team\_max\_score}$ is the player's maximum score up to time t, Y is the vector of the 10 highest scores on the leaderboard up to time t, $\mathit{top}1$ is an indicator for whether the player is in the first position of the leaderboard, $\mathit{top}3$ is an indicator for whether the player is in positions 2 or 3 of the leaderboard, and $\mathit{top}10$ is an indicator for whether the player is in positions 4 to 10 of the leaderboard.

The likelihood is constructed using tuples $\{(s_i, t_i, t'_i, \theta_i)\}_{i\in N}$, where i is a submission, s_i is the vector of state variables at the moment of making the submission, t_i is the submission time, and t'_i is the arrival time of the next submission, which may or may not be observed, and θ_i is player type. If the next submission is observed, then $t_i < t'_i \leq T$, if not, $t'_i > T$. If the new submission arrives at $t'_i \leq T$, then the player must have chosen to make a new submission at t_i, and the likelihood of the observation $(s_i, t_i, t'_i, \theta_i)$ is given by $l(s_i, t_i, t'_i, \theta_i) = \Pr(\text{play}|s_i, t_i, \theta_i) \cdot \lambda e^{(-\lambda(t'_i - t_i))}$, where $\lambda e^{(-\lambda(t'_i - t_i))}$ is the density of the submission arrival time. If we do not observe a new submission after the player's decision at time t (i.e., $t'_i > T$), then the likelihood of $(s_i, t_i, t'_i > T, \theta_i)$ is given by $l(s_i, t_i, t'_i > T, \theta_i) = \Pr(\text{play}|s_i, t_i, \theta_i) \cdot e^{(-\lambda(T - t_i))} + 1 - \Pr(\text{play}|s_i, t_i, \theta_i)$, which considers both the events of i) the player choosing to make a new submission at t_i and the submission arriving after the end of the contest; and ii) the event of the player choosing not to make a new submission.

B.5.2 Benefits of building a new submission

For every (s, t, θ) in the data, we simulate $NS = 200$ continuation histories of the game under two cases: the player chooses to build a new submissions (d = Play) and the player chooses not to build a new submission (d = Not Play). Histories are simulated using estimates for all primitives of the model (except for the distribution of costs) and the CCPs. For every simulated history, we compute the payoff of the player at the end of the game. For simplicity, we assume that every competition is a winners-take-all contest. The simulated benefit of action d given state variables (s, t, θ) is then the average payoff at the end of the game across all simulated histories that follow decision d.

C Description of the Experiment

Description of the Competition

A large restaurant chain owns restaurants located along major highways. The average revenue of a restaurant located at distance x from the highway is $R(x)$. For simplicity, the variable distance to the highway is normalized to be in the interval [1,2]. The function $R(x)$ is unknown. The goal of this competition is to predict the value of $R(x)$

for several values of distances to the highway. Currently, the restaurant chain is located at 40 different locations. You will have access to

$$\{(x_i, R(x_i))\}_{i=1}^{30},$$

i.e., the distance to the highway and average revenue for 30 of these restaurants. Using these data, you must submit a prediction of average revenue for the remaining 10 restaurants, using their distances to the highway.

You will find the necessary datasets in the Data tab. You can send up to 10 different submission each day until the end of the competition. The deadline of the competition is Sunday April 15th at 23:59:59.

Evaluation

We will compare the actual revenue and the revenue predictions for each value

$$(x_j)_{j=31}^{40}.$$

The score will be calculated according to the Root Mean Square Deviation:

$$\text{RMSD} = \sqrt{\frac{\sum_{j=31}^{40}(\hat{R}(x_j) - R(x_j))^2}{10}},$$

which is a measure of the distance between your predictions and the actual values $R(x)$. The deadline of the competition is Sunday April 15th at 23:59:59.

Note. Following the convention used throughout the paper, we multiplied the $RSMD$ scores by minus one to make the winning score maximize private score in the competition.

Description of the Data

The goal of this competition is to predict the value of $R(x)$ for a number of values of distance to the highway. The csv file "train" contains data on the distance to the highway and average revenue for 30 restaurants

$$\{(x_i, R(x_i))\}_{i=1}^{30},$$

You can use these data to create predictions of average revenue for the remaining 10 restaurants. For these 10 restaurants you only observe their distances to the highway in the csv file "test." You can find an example of how your submission must look like in the csv file "sample_submission."

File descriptions:

- **train.csv** - the training set
- **test.csv** - the test set
- **sample_submission.csv**- an example of a submission file in the correct format

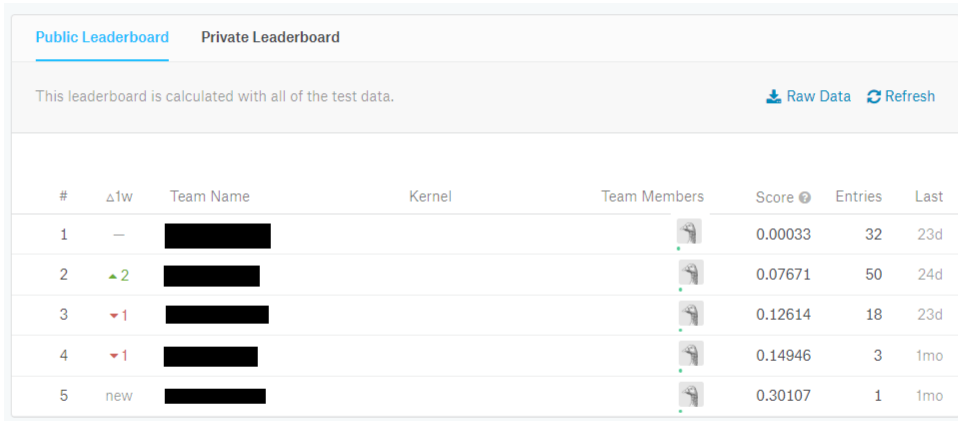
Submission File:

The submission file must be in csv format. For every distance to the highway of the 10 restaurants, your submission files should contain two columns: distance to the highway (x) and average revenue prediction (R). The file should contain a header and have the following format:

x	R
1.047579	34.43375
1.926801	36.83077
etc.	

A correct submission must be a csv file with one row of headers and 10 rows of numerical data, as displayed above. To ensure that you are uploading your predictions in the correct format, we recommend that you upload your predictions by editing the sample submission file. There is a limit of 10 submissions per day.

Figure A.3 shows a screenshot of the leaderboard in one of our student competitions hosted in Kaggle.



Public Leaderboard   Private Leaderboard

This leaderboard is calculated with all of the test data.   Raw Data   Refresh

#	△1w	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	[redacted]			0.00033	32	23d
2	▲2	[redacted]			0.07671	50	24d
3	▼1	[redacted]			0.12614	18	23d
4	▼1	[redacted]			0.14946	3	1mo
5	new	[redacted]			0.30107	1	1mo

Figure A.3: Snapshot of the leaderboard in one of our competitions with a leaderboard. Names are hidden for privacy reasons.