Dynamic Tournament Design: An Application to Prediction Contests*

Jorge Lemus† Guillermo Marshall‡

May 24, 2017

[PRELIMINARY AND INCOMPLETE]

Abstract

We build a tractable structural model to study the dynamic participation incentives in online competitions. We estimate the model using publicly available data from 17 prediction contests hosted by the website Kaggle.com. We study how participants' incentives are shaped by the contest design in a series of counterfactual exercises. Specifically, we study how disclosing a leaderboard, awarding different number of prizes, and capping the number of participants affect the participants' incentives to make submissions throughout the contest as well as the contest outcomes.

Keywords: Dynamic contest, contest design, prediction, Kaggle, big data

*We thank seminar and conference participants at International Industrial Organization Conference, Rob Porter Conference (Northwestern), Triangle Microeconomics Conference (UNC), and University of Georgia for helpful comments. The usual disclaimer applies.

†University of Illinois at Urbana-Champaign, Department of Economics; jalemus@illinois.edu

‡University of Illinois at Urbana-Champaign, Department of Economics; gmarshll@illinois.edu

1 Introduction

Online tournaments are widely used by government agencies and private companies to outsource innovative approaches to solve problems. Since 2010, U.S. government agencies have sponsored over 730 competitions that have awarded over $250 million in prizes to procure software, ideas, or designs through the website www.challenge.gov. For instance, DARPA launched a competition to accurately predict cases of chikungunya virus, awarding a total reward of $500,000.[1] Apart from government sponsored contests, several two-sided platforms that match companies' problems and data scientists have recently emerged.[2] We collect publicly available data from one of these platforms and build a structural model to study economic incentives in a prominent class of online competitions called *prediction contests*—contests where the goal is to procure a model that delivers accurate predictions of a random variable. In a series of counterfactual exercises, we study how contest design shapes incentives and impacts contest outcomes. In the competitions that we analyze, the designer chooses to award multiple prizes, does not restrict the number of players, allows players to make multiple submissions over time, and discloses a real-time noisy ranking of participants based on their performance in the contest.[3] We study how dynamic participation and contest outcomes would change when varying the information about contestants' performance that is made public,[4] if there was a single prize, or if entry was restricted.

The "Big Data" revolution,[5] facilitated by the advances in computer power and storage technology, has given firms and individuals the ability to create and store greater amounts of information. Large datasets offer the possibility to analyze complex problems, and online competitions offer an alternative to seek solutions to these problems. The appeal of online competitions is that they attract participants with different abilities and expertise and may facilitate the procurement of solutions that otherwise would had never been found (Lakhani et al., 2013). By developing an accurate predictive

[1] http://www.darpa.mil/news-events/2015-05-27

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

[3] The ranking is noisy in the sense that at any moment of time the ranking that would be used to determine the winner of the competition is not the same as the ranking that is disclosed to participants. Both rankings, however, are highly correlated. We provide more details below.

[4] One of the designs that we study is a "blind tournament," as in Taylor (1995).

[5] http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

algorithm from a large dataset of past observations, it is possible to diagnose diseases early on, to prevent epidemics from spreading in a population, or to manage inventory when there is fluctuating demand. Hence, it is important to understand how the design of online competitions affect participants' incentives and contest outcomes.

We use public information on contests hosted by Kaggle,[6] a company dedicated to host prediction contests. Through this platform, different companies have sponsored over 200 competitions that have awarded over $5 million dollars in prizes. For instance, EMI sponsored a $10,000 contest to predict if listeners would like a new song; IEEE sponsored a $60,000 contest to diagnose schizophrenia using multimodal features from MRI scans; The National Data Science Bowl sponsored a $175,000 contest to predict the ocean's health by identifying plankton species from multiple images. In all of these contests, participants are provided with a training dataset for them to develop a prediction algorithm, and then their algorithms are evaluated based on how well they perform out of sample. The contests that we analyze have an *objective* evaluation criterion that is disclosed to participants at the beginning of the contest.[7] The objective evaluation criterion in prediction contests is in contrast to the evaluation criteria in ideation contests (Huang et al., 2014; Kireyev, 2016), innovation contests (Boudreau et al., 2016), design contests (Gross, 2015), or labor promotions (Lazear and Rosen, 1979; Baker et al., 1988), where evaluation (or some part of it) has a *subjective* component.

Aside from the training dataset, participants of Kaggle contests are also provided with a test dataset. Unlike the training dataset which includes both an outcome variable and covariates, the test dataset only includes covariates. The test dataset is used to evaluate the out of sample performance of the participants' prediction algorithms. The performance of an algorithm is quantified using the contest's evaluation criterion. For an algorithm to compete for the prize, participants must submit outcome variable

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

[7] For example, in the ocean's health competition, the participants submitted their predictions (p_{ij}) and were evaluated according to the rule

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij}).$$

The winner was the submission with the lowest score. For more details, visit: https://www.kaggle.com/c/datasciencebowl/details/evaluation.

predictions for each observation in the test dataset (i.e., make a submission). Kaggle divides the test data in two subsets without informing participants how the division is made.[8] The first test data subset is used to generate a *public score*, which is posted in real-time on a public leaderboard that can be observed by any Internet user. The predictions on this second test data subset generate a *private score*, which is never observed by players before the end of the contest. The winner of a Kaggle competition is the player with the maximum private score. In our data, the correlation between public and private is 0.99, but only 76 percent of the contest winners finish in the top 3 of the public leaderboard.

Our paper contributes to the fairly recent literature of empirical studies of contests. The main novelty of our paper is to build a tractable model to study incentives in prediction contests where players are allowed to make multiple submissions over time and information is disclosed throughout the contest. Modeling a dynamic prediction contest poses various economic questions and technical challenges. First, although there is a real-time leaderboard, contestants are always uncertain of their actual position in the contest. In other words, the contestants only receive a noisy signal of their current position in the contest. Despite the fact that private and public scores are highly correlated, the rankings under the public and private scores vary. This feature may create an encourage or discouragement effect. Second, from a contest design perspective, the decision of disclosing the public ranking may effect the decision to continue participating in the contest. Third, on the technical side, there is a large number of heterogeneous participants sending thousands of submissions. An analytic solution for a dynamic model with heterogeneous and fully-rational players is cumbersome. In fact, given that participants are unsure of their position in the leaderboard, they need to keep track of the whole public history to in order to compute the payoff of sending an extra submission. The dimensionality of the public information released over time makes the state state space become computationally intractable. To deal with this problem, we assume players are *small*—i.e., they do not take into account the effect of their actions on other player's strategies—and also limit the amount of information that players believe is relevant for predicting their chances of winning the contest.

[8]More specifically, Kaggle does not inform participants which observations in the test data correspond to each subset.

In the model, participants are heterogeneous in their ability to produce algorithms that achieve high scores. In practice, this means that participants draw submissions from different score distributions. We assume that once players enter a contest, they work on at most one submission at a time. Conditional on choosing to build an algorithm (i.e., make a submission), we assume that finishing the submission takes the player t units of time, where t is a random variable. Once the submission is finished, the player chooses whether to build a new submission or to exit the contest. When deciding whether to build a new submission, the player compares the benefit of building a new submission (i.e., how the submission increases the players' expected payoff) with the cost of making the submission, which is a random variable. In computing the benefit of making a submission, the player considers her chances of building a winning submission given the current public leaderboard scores and also considering the fact that other players will make more submissions in the remaining contest time, and that those submissions will lower the player's chance of winning the contest.

Our results show that contest design matters and there is no one-size-fits-all policy prescription. Our counterfactual simulations show that there is heterogeneity in the response of both participation and contest outcomes to different contest designs. We present our results in terms of how contest design impacts the number of submissions—our measure of participation—because an increase in the number of submissions generally leads to an increase in the maximum score. We find that manipulating the amount of information disclosed to participants has economically significant effects on the number of submissions. If the contest designer hid the leaderboard—that is, if the contest designer did not provide public information about contestants' performance—the number of submissions would increase on average by 11 percent. Increasing the correlation between the private and public scores—a policy that increases the amount of information about contestants' performance that is made public—would decrease the number of submissions by 2.7 percent. Allocating a single prize rather than several prizes has a small and insignificant effect on the number of submissions. Limiting the number of players that participate in a contest has a positive effect on the individual dynamic incentives to participate but also a direct negative effect on participation. We find that when the number of participants is reduced by 10 percent in each contest, the increase in participation from each individual does not compensate for the reduction in the number of players and the total number of submissions falls by an average of 8.7 percent. In summary, these results suggest that the most effective tool for the contest designer to increase the average number of submissions is manipulating information disclosure.

Finally, it is worth mentioning that participation in these online competitions is not solely driven by the monetary prize. Participants can learn new skills by working with new types of problems and by sharing their ideas with other researchers. Similarly to software engineers contributing on open-source software (Lerner and Tirole, 2002), performing well in a data-science competition is a signal of quality to potential future employers. Hence, our cost estimates of the cost of making a submission also capture the non-monetary rewards of participating.

In the next subsection we review the relevant literature. In Section 2 we present descriptives that motivate the modeling choices. Section 3 introduces the empirical model and Section 4 discusses estimation and presents the model estimates. Section 5 presents counterfactual exercises to grasp the effect of various alternative contest designs. Finally, in Section 6, we summarize our results and provide further discussion.

1.1 Related Literature

Contests are a popular open innovation mechanism (Chesbrough et al., 2006). An extensive theoretical literature analyzes how the design of a contest—the number of participants, the number of prizes, and the disclosure of information—affects the incentives to participate and the quality of the best submission.

Our counterfactuals analyze how the number of participants and the distribution of prizes affects incentives to participate. Taylor (1995) and Fullerton and McAfee (1999) show that restricting the number of competitors in winner-takes-all tournaments increases equilibrium outcomes. Intuitively, with a large number of participants players have less incentives to exert costly effort because they have a smaller chance of winning. Moldovanu and Sela (2001) show that, when players have different abilities, the optimal number of prizes depends on the shape of the cost of effort function. Che and Gale (2003), in a model where the designer chooses a menu of prizes and the number of participants, find that one prize and two participants is optimal.

The conclusion of most of these papers is that limiting the number of competitor is optimal, although in practice the contest designer typically encourages free-entry participation, as in the contests we analyze. Terwiesch and Xu (2008) incorporates diversity into the preferences of the contest designer counteracting the effect of participation on effort. Boudreau et al. (2011) tests this theory and show that both of these forces interact. They find that more competitors shifts the distribution of outcomes to the left, but the maximum outcome is higher. Boudreau et al. (2016) find that competitors of different ability respond differently to the number of participants. Their empirical finding shows that low-ability players are almost unresponsive to the total number of participants, while medium-ability (high-ability) players decrease (increase) their effort when there are more rivals.

The effect of information disclosure on incentives in dynamic tournament has also been studied. If players observe a noisy signal of the effort exerted by their rivals, releasing information creates dynamic asymmetries between the contestants. Aoyagi (2010) explores a dynamic tournament and compares the effort provision by agents under full disclosure of information (i.e., players observe their relative position) versus no information disclosure. Which one of these two disclosure policies dominates depends on the shape of the cost of effort function. This setting is further explored by Ederer (2010), who adds private information, and by Klein and Schmutzler (2016), who add design features such as the allocation of prizes and different forms of evaluating performance.

The environment we study is dynamic and each contestant can submit multiple submissions conditional on the history of the contest. Konrad (2012) reviews the literature on dynamic contests. Dynamic contest design has been recently studied by Halac et al. (2014), Bimpikis et al. (2014), and Benkert and Letina (2016). Takahashi (2015) empirically estimates a war of attrition.

There is a growing empirical literature analyzing different classes of contests. Gross (2015) studies how the number of participants changes incentives to create 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 and quality of outcomes. In these 'ideation contests,' the evaluation is subjective. Bockstedt et al. (2016) descriptively study the effect of full feedback on participation. Finally, Huang et al. (2014) estimates a dynamic structural model to study individual behavior and outcomes in a platform where individuals can contribute ideas and some of them are implemented.

Another strand of the literature that relates to our research is the question of why people spend time and effort participating in contests where there monetary reward is small or even non-existent. Lerner and Tirole (2002) study why people contribute to open software and they argue that good quality contributions are a signal of ability to potential employers. Related to this point, participants may have private information on the actual benefit from winning the contest. Chawla et al. (2015) study optimal contest design of crowd-sourcing contests when participants have private information about their value of winning. Alternatively, people may just enjoy participating in a contest because it gives them social status (Moldovanu et al., 2007).

Apart from prizes, number of competitors, and feedback, there are other design tools. Megidish and Sela (2013) consider contests in which participants must exert some (exogenous) minimal effort and show that awarding a single prize is dominated by giving each participant an equal share of prize when the exogenous threshold for participation is high. Moldovanu and Sela (2006) show that for a large number of competitors it is optimal to split them in two divisions. In the first round participants compete within each of these divisions, and in the second round the winners of each division compete to determine the final winner.

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), there are fewer papers analyzing behavior in contests. We see our contribution as one of the first papers to study contest design in a dynamic setting with objective evaluations.

2 Background, Data, and Motivating Facts

2.1 Background and Data

We use publicly available information on contests hosted by Kaggle.[9] The dataset contains several types of competitions. Featured competitions are public competitions to solve commercial problems. The winners grant the sponsor a non-exclusive license to their submissions in exchange for a monetary award. These competitions represent about 75 percent of the competitions in the data. Research competitions (16 percent of the competitions in the data) are public competitions with the goal of providing a public good. Prizes for research competitions include monetary awards, conference invitations, and publications in peer-reviewed journals. The winners of research competitions must release their solutions open source. Other contest categories include competitions for recruiting (0.32 percent of the competitions in our data), competitions for data visualization (2.25 percent of the competitions in the data), and competitions for fun (4.5 percent of the competitions in the data).

We work with a subset of 55 featured competitions, all of which offered a monetary prize of at least \$1,000 and received at least 1,000 submissions. These contests have the feature that the winning submission either maximizes or minimizes a well-defined rule, and have an informative public score leaderboard. In these competitions, there was an average of 1,755 teams per contest, competing for rewards that ranged from \$1,000 to \$500,000 and averaged \$30,642. On average, 15,169 submissions were made per contest. The characteristics of a partial list of competitions are summarized in Table 1 (see Table A.1 in the Online Appendix for the full list). All of these competitions, with the exception of the Heritage Health Prize, granted prizes to the top three scores.[10] For example, in the Coupon Purchase Prediction competition, the three submissions with the highest scores were awarded \$30,000, \$15,000, and \$5,000, respectively.

As mentioned in the Introduction, one interesting feature of Kaggle is its rules to determine the contest winner. There is a large dataset partitioned into three subsamples.

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

[10] The following contests also granted a prize to the fourth position: Don't Get Kicked!, Springleaf Marketing Response, KDD Cup 2013 - Author Disambiguation Challenge (Track 2).

Name of the Competition	Total Reward	Number of Submissions	Teams	Start Date	Deadline
Heritage Health Prize	500,000	25,316	1,353	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	25,195	952	04/10/2014	07/14/2014
Liberty Mutual Group - Fire Peril Loss Cost	25,000	14,812	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	23,244	1,326	05/11/2015	07/06/2015
Caterpillar Tube Pricing	30,000	26,360	1,323	06/29/2015	08/31/2015
Liberty Mutual Group: Property Inspection Prediction	25,000	45,875	2,236	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	39,444	2,226	08/14/2015	10/19/2015
Homesite Quote Conversion	20,000	36,368	1,764	11/09/2015	02/08/2016
Prudential Life Insurance Assessment	30,000	45,490	2,619	11/23/2015	02/15/2016
Santander Customer Satisfaction	60,000	93,559	5,123	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.

The first subsample can be used by the contestants to develop their predictions. This first subsample provides both outcome variables and covariates. The second and third subsamples are the test data, and these subsamples are used for evaluation. Both test subsamples are provided to the players as a single dataset and only include covariates (i.e., no outcome variables). Kaggle computes the *public score* and *private score* by evaluating how well a player's submission predicts the outcome variables in the second and third subsample, respectively. For example, in the Heritage Health Prize, the test data was divided into a 30 percent subsample to compute the public scores and a 70 percent subsample to compute the private scores. Kaggle does not disclose what part of the test data are used to compute the public and private scores.

Kaggle displays, in real-time, a public leaderboard which contains all the public scores of the submissions of all participants. Since these public scores are calculated by only using part of the test dataset (e.g., 30 percent in the Heritage Health Prize competition), the final standings may be different than the ones displayed in the public leaderboard. Although the correlation between public and private scores is very high in our sample (the coefficient of correlation is 0.99), the ranking in the public leaderboard is not necessarily equal to the ranking in the private leaderboard. Hence, the public leaderboard provides informative yet noisy signals on the performance of all players throughout the contest. To illustrate this noise, consider the winner of each of the 55 competitions that we analyze—i.e., the owner of the submission with the highest private score (see Table A.2 in the Online Appendix). In 27 out of 55 competitions (49 percent), the winner of the contest was ranked number one in the final public leaderboard, and in 42 out of 55 competitions (76 percent) the winner was within the top three of the final public leaderboard. That is, while the public leaderboard is informative about the players' payoff-relevant performance, the co-existence of the public and private scores creates uncertainty about the true standing of a player in the competition.

2.2 Motivating Facts

We present a series of empirical facts that will guide our modeling choices. For each contest, we observe information on all submissions including the time of submission, the team identity, and both the public and private scores of the submission. With this information, we can reconstruct both the public and private leaderboard at every instant of time. Throughout the paper, we normalize the contest length to one, implying that the submission times always lie on the unit interval. Also, we normalize the total prize to one and transform the public and private scores to be contained in the unit interval by applying the transformation $(x_i - \min_j x_j)/(\max_j x_j - \min_j x_j)$ to each observation of variable x. These normalizations allow us to make meaningful comparisons across contests.

We start by examining some summary statistics. Table 2 (Panel A) shows that the (transformed) public and private score take an average value of 0.88, with a standard deviation of 0.2. The average submission time is when 60 percent of the contest time has elapsed, and two consecutive submissions by the same team are spaced in time by on average of 2 percent of the contest duration. Panel B shows that teams on average send 16.38 submissions per contest, with some teams sending as many as several hundred. Lastly, 93 percent of the teams are composed by a single member, leading to an average team size of 1.13.[11]

[11] Table A.3 in the Online Appendix shows that 72 percent of users participate in a single contest,

Panel A: Overall summary statistics

	N	Mean	St. Deviation	Min	Max
Public score	834,301	0.88	0.20	0.00	1.00
Private score	834,301	0.88	0.20	0.00	1.00
Submission time	834,301	0.60	0.29	0.00	1.00
Time between submissions	783,362	0.02	0.05	0.00	1.00

Panel B: Team-level statistics

	N	Mean	St. Deviation	Min	Max
Number of submissions	50,937	16.38	29.13	1	671
Number of members	50,937	1.13	0.61	1	40

Table 2: Summary Statistics

Note: An observation in Panel A is a submission; an observation is a team–competition combination in Panel B. Scores and time are rescaled to be contained in the unit interval (i.e., we apply the transformation $(x_i - \min_j x_j)/(\max_j x_j - \min_j x_j)$ to each observation of variable x). Time between submissions is the time between two consecutive submissions by the same team.

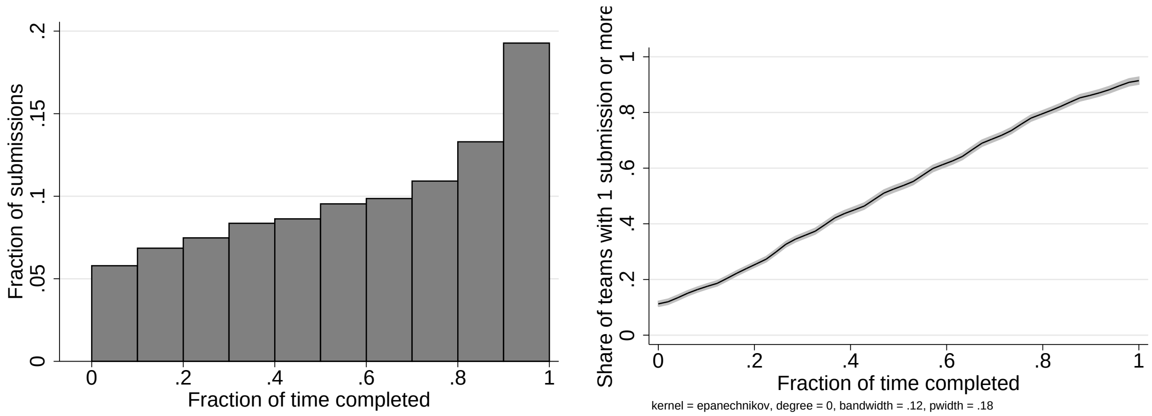
Observation 1. *Most teams are composed by a single member.*

Figure 1 shows how the number of submissions and teams evolve over time. Panel 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 being submitted when 10 percent of the contest time has elapsed. Panel 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. *The rate of entry of new teams is constant throughout the contest duration.*

To understand whether teams become more or less productive as time elapses, we examine the time between submissions at the team level. Figure 2 (Panel A) illustrates

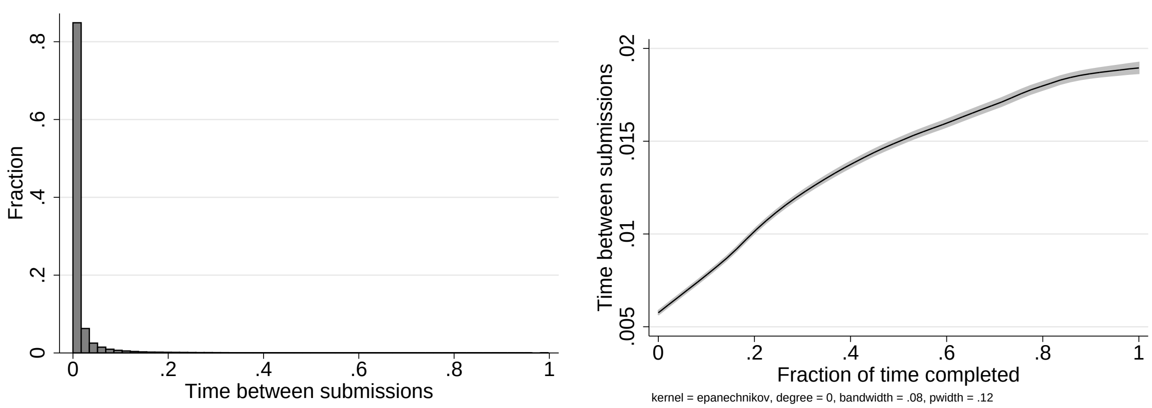
suggesting that most players are one-off participants.



(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.



(a) **(b)**

Figure 2: Time Between Submissions

Note: An observation is a submission. Panel (a) shows the distribution of time between two submissions. Panel (b) shows a local polynomial regression of the time between submissions as a function of time.

the time between two consecutive submissions by the same team. On average, teams take 2 percent of the contest time to send two consecutive submissions. Panel B shows a local polynomial regression for the average time between submission as a function of time. The figure shows that the average time between submissions increases over time, suggesting that either that times are experimenting when they enter the contest or that finding new ideas 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 teams making submissions at a faster pace over time, but simply because there are more active teams at the end of the contest and potentially greater incentives to play.

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

	(1)	(2)
	Public Score	
Second 25 percent of submissions		0.0445***
		(0.0004)
Third 25 percent of submissions		0.0624***
		(0.0004)
Last 25 percent of submissions		0.0744***
		(0.0004)
Competition × Team FE	Yes	Yes
Observations	826,310	826,310
R^2	0.696	0.715

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. Second 25 percent of submissions is an indicator variable for whether a submission is within the second 25 percent of submissions of a team, where submissions are sorted by submission time. The other indicators are defined analogously.

Figure 3 shows the joint distribution of public and private scores for all submissions. The coefficient of correlation between both scores is 0.99.[12] Table 3 decomposes the variance of public scores. In column 1, we find that 70 percent of the variation in

[12]Notice the cluster of points around (0.3,0.9). These scores have a low private score (around 0.3) but a high public score. This is an example of overfitting: submissions that deliver a large public score but they are poor out-of-sample predictors (i.e., not robust submissions).

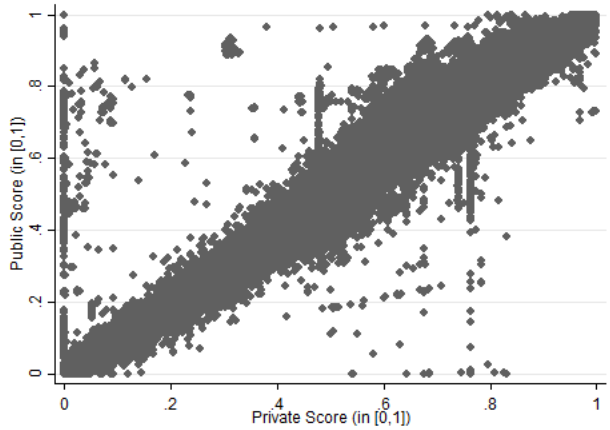


Figure 3: Correlation Between Public and Private Scores

Note: An observation is a submission. The private and public scores of each submission are normalized to range between 0 and 1.

public score is between-team variation, suggesting that teams differ systematically in the scores that they achieve. In column 2, we allow for dummies that distinguish each team's submissions by whether they were early or late submissions (with respect to each team's set of submissions). This distinction allows us to measure whether later submissions made by a team achieved systematically greater scores than earlier submissions. The table shows that there are within-team improvements over the course of the contest, although those improvements only explain an additional 1.9 percent of the overall public score variance. In the model, we will capture these cross-team differences by allowing for the teams to systematically differ in their ability to produce high scores. We leave within-team dynamics and learning for future research.

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

With respect to how the public leaderboard shapes behavior, Table 4 suggests that teams drop out of the competition when they start falling behind in the public score leaderboard. In the table, we compare how the timing of a team's last submission varies with the score gap between the maximum public score and their best public score up to that moment. A one standard deviation increase in a team's deviation from the maximum public score is associated with a team submitting its final submission (0.03× total contest time) to (0.08× total contest time) sooner. That is, teams that are

	(1)	(2)
	Timing of last submission	
Deviation from max	-0.0327***	-0.0782***
public score (standardized)	(0.0012)	(0.0018)
Competition FE	Yes	Yes
Weights	No	Yes
Observations	50,937	50,937
R^2	0.050	0.065

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

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Timing of last submission is measured relative to the total contest time (i.e., it ranges between 0 and 1). Deviation from max public score is defined as the competition wide maximum public score at the time of the submission minus the submitting team's maximum public score at the time of the submission. We then standardize this variable using its competition-level standard deviation. Column 2 weighs observations by the total number of submissions made by each team.

	(1)	(2)
	Number of submissions	log(Number of submissions)
After disruptive submission	-0.6070**	-0.0748***
	(0.2741)	(0.0247)
Competition FE	Yes	Yes
Observations	2,531	2,531
R^2	0.755	0.764

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 defined as submissions that increase the maximum public score by at least 1 percent. Number of submissions is the number of submissions in time intervals of length 0.001. The regressions restrict the sample to periods that are within 0.05 time units of the disruptive submission. Both specifications control for time and time squared.

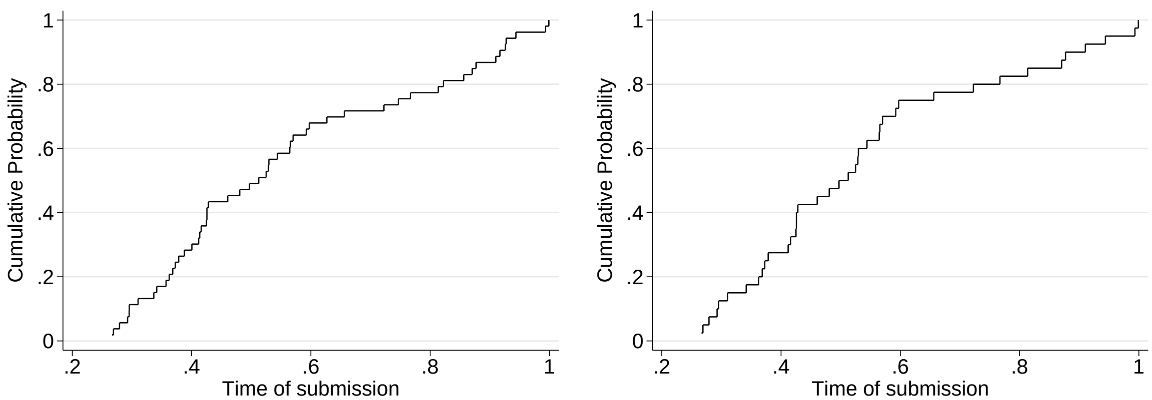
lagging behind seem to suffer a discouragement effect and quit the competition. This exercise sheds light on how information disclosure may affect participation incentives throughout the competition.

We also analyze how the public leaderboard shapes incentives to participate in Table 5. In this table, we analyze how the rate of arrival of submissions changes when the maximum public score jumps by a significant margin. Whenever a submission increases the maximum public score by a sufficient amount (e.g., 1 percent for our analysis in Table 5), we call the submission *disruptive* (see Figure A.1 in the Online Appendix for an example). Only 0.05 and 0.04 percent of submissions increased the maximum public score by 0.5 and 1 percent, respectively. To measure how the rate of arrival of submission changes with a disruptive submission, we first partition time in intervals of length 0.001 and compute the number of submissions that arrive 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. Table 5 shows that the number of submissions decreases immediately after the disruptive submission by an average of 7.5 percent. We take this as further evidence of both the discouragement effect and how the public leaderboard affects the behavior of participants.

Observation 5. *The public leaderboard shapes participation incentives.*

With respect to the timing of submissions that disrupt the leaderboard, Figure 4 plots the timing of submissions that increased the maximum public score by at least 0.5 percent (Panel A) and 1 percent in (Panel B). In the figure we restrict attention to submissions that were made when at least 25 percent of the contest time had elapsed because score processes are noisier earlier in contests. The figure suggests that disruptive submissions arrive uniformly over time. This pattern suggests that teams are not being strategic about the submission time of solutions that they believe will drastically change the public leaderboard. This may be driven by the fact that teams only learn about the out-of-sample performance of a submission after Kaggle has evaluated the submission. That is, before making the submission, the teams can only evaluate the solution using the training data, which may not be informative about the solution's out-of-sample performance.

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



(a) Increase greater than 0.5 percent

(b) Increase greater than 1 percent

Figure 4: 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 x percent. The figure plots submissions that were made when at least 25 percent of the contest time had elapsed.

3 Empirical Model

We consider a contest with a well-defined evaluation system. The length of the contest, $T > 0$, is normalized to $T = 1$. Players are ranked at the end of the contest and the first j-th players in the ranking receive prizes of value $V_1 \geq ... \geq V_j$.

There is a fixed supply of N players of heterogeneous ability, which is captured by the set of types $\Theta = \{\theta_1, ..., \theta_p\}$.[13] We allow for heterogeneous players because Observation 4 suggests heterogeneity among players. The distribution of types, $\kappa(\theta_k) = \Pr(\theta = \theta_k)$, is known by all players. We assume that players enter the contest at a random

[13]Following Observation 1, we disregard any team behavior and treat each participant as a single player.

time τ_{entry} distributed according to an exponential distribution of parameter $\mu > 0$.[14] The constant rate of entry described by Observation 3 justifies this assumption. The empirical evidence does not strongly suggest that players strategically choose the time of entry, but rather they enter at a random time, possibly related to idiosyncratic shocks such as when they find out about the contest. In our model, although players can send multiple submissions throughout the contest, they can work at most on one submission at the time. Working on a submission takes a random time τ distributed according to an exponential distribution of parameter λ. In general, the parameter λ could be a function of the type of the player, the number of submissions already sent, or the time that is left in the contest. For simplicity, however, we assume that λ is a constant. Observation 3 and Figure 1 motivate this assumption. The behavior shown in Figure 2, however, suggests a richer dynamic (e.g., experimentation) which we do not model and leave for future work. An unobserved element of the model is the cost of building submissions, $c \sim K(\sigma)$, where σ is a parameter to be estimated. The decision of a player when presented with an opportunity to play is to either build a submission and continue playing or to quit the contest.

The evaluation of a submission is based on the solution sent by a player and a test dataset d. Each pair $(\text{solution}, d)$ maps uniquely into a score through a well-defined formula. Motivated by the evaluation system used in practice, we consider two test datasets, d_1 and d_2, which define two scores: the *public score*, computed using the solution submitted by the player and test dataset d_1; and the *private score*, computed using the solution submitted by the player and test dataset d_2. Instead of modeling the solution submitted by the players, we model the score of a submission as a random variable. A player of type θ draws a public-private score pair $(p_{\text{public},\theta}, p_{\text{private},\theta})$ from a joint distribution $H_\theta([0,1]^2)$. Players know the joint distribution H_θ, but they do not observe the realization $(p_{\text{public},\theta}, p_{\text{private},\theta})$. This pair of scores is *private information* of the contest designer. In the baseline case, the contest designer discloses, in real time, only the public score $p_{\text{public},\theta}$ but not the private score $p_{\text{private},\theta}$. The final ranking, however, is constructed with the private scores.[15] Figure 3 shows that public and private

[14] When players enter the competition, we assume they get a free submission. This assumption is standard in the search literature (Diamond, 1971).

[15] Although players are allowed to send multiple submissions—and each player sends about 20 submissions on average—the final ranking is computed by using at most two submissions by each player.

scores are highly correlated.[16]

The contest designer releases, in real time, the public scores and the identity of the players that obtained those scores. The collection of pairs (identity, score) from the beginning of the contest to instant t conforms the *public leaderboard*, denoted by $\mathcal{L}_t = \{(\text{identity, score})_j\}_{j=1}^{J_t}$, where J_t is the total number of submissions up to time t. Conditional on the terminal public history $\mathcal{L}_T$, each player i is able to compute $p_{\ell,i}^{\text{final}} = \Pr(i\text{'s private ranking is } \ell | \mathcal{L}_T)$, which is the probability of ranking in position ℓ in the private leaderboard, conditional on the final public leaderboard $\mathcal{L}_T$.

A model with fully-rational players is challenging for several reasons. First, it is possible that $p_{1,i}^{\text{final}} > 0$ even if player i is ranked last in the public leaderboard (though unlikely). That is, every player that has participated in the contest has a non-zero chance of winning, regardless of their position in the public leaderboard. Hence, players must use all the available information in the public leaderboard every time they decide whether to play or not. Keeping track of the complete history of submissions, with over 15,000 submissions in each competition, is computationally intractable.[17] In contrast to a dynamic environment in which players perfectly observe their relative position, the public leaderboard is just a noisy signal of the actual position of the players in the contests. Without noise, i.e., the player with the highest *public* score at the terminal history is the winner, players only need to keep track of the current highest public score to make their investment decision, which leads to a low-dimensional state space. In our setting, however, the state space is large because the relevant public history is not summarized by a single number. To overcome this computational difficulty, we assume that $p_{\ell,i}^{\text{final}} > 0$ for $\ell = 1, 2, 3$ if and only if player i is among the three highest scores in the final public leaderboard. In other words, we assume the final three highest private scores are a permutation of final three highest public scores. This assumption is motivated by Table A.2 in the Online Appendix, because in 76 percent of the contests

---

Players are given the option to select which two submissions will be evaluated by Kaggle to determine the final standings. About 50 percent of the players do not make a choice, in which case Kaggle picks the two largest public scores. Out of the 50 percent remaining that indeed choose, 70 percent choose the two scores with the highest public score.

[16]This assumption can be easily relaxed with more computational power (time).

[17]For example, if we partition the set of public scores into 100 values, with 15,000 submissions the number of possible terminal histories is of the order of 2^{300}.

we study, the winner is among the three highest public scores.

Small and Myopic Players

There are at least 15,000 submissions and thousands of players on average in each contest. Fully rational players would take into account the effect of their submissions on the strategy of the rival players. However, solving analytically a model with fully rational players turns out to be extremely challenging. As a simplification, we assume that players are *small*, i.e., they do not consider how their actions affect the incentives to play of other players. This price-taking-like assumption is not unreasonable for our application. Observation 6, partially justify this assumption.

Additionally to assuming that players are small, we make another simplification which we will later relax. We assume that when players make the decision to play or to quit, they expect more submissions in the future by rival players but not by themselves. In other words, *myopic players* think this current opportunity to play will be their last one. It is worth noting that under this assumption players might play multiple times, however they *think* they will never have a future opportunity to play or in case they do they will choose not to play. This means that myopic players are not sequentially rational. Later on, in subsubsection 3.1.2, we analyze a model with sequentially rational players.

We start the analysis under the small and myopic players assumptions, i.e., players that do not consider the effect of their actions on the other players' decision to play and also do not account for the possibility of playing again in the future.

State Space and Incentives to Play

The relevant state space is defined by three sets. First, we define the set of (sorted) vectors of the three largest public scores, $\mathcal{Y} = \{y = (y_1, y_2, y_3) \in [0,1]^3 \ : \ y_1 \geq y_2 \geq y_3\}$. Second, we define $\mathcal{R}_S = \{\emptyset, 1, 2, 3, (1,2), (1,3), (2,3)\}$ to be the set of *score ownership*. The final set is $T = [0,1]$ which represents the contest's time. Notice that $y \in \mathcal{Y}$ and $t \in T$ are public information common to all players. Under the small-player assumption, the relevant state for each player is characterized by $s = (t, r_i, y) \in \mathcal{S} \equiv T \times \mathcal{R}_S \times \mathcal{Y}$. To be precise, $s = (t, r_i, y) \in \mathcal{S}$ means that at time t player i owns the components of vector y indicated by r. For example, $(t, (1,3), (0.6, 0.25, 0.1))$ means that at time t, the player components one and three in vector y, i.e., the player owns two out of the three highest public scores: 0.6 and 0.1.

The small-player assumption reduces the dimensionality of the state space, because players care only about the three highest public scores and which one of them they own. Also, although they do not observe the private scores, they are able to compute the conditional distribution of private scores given the set of public scores. Because prizes are allocated at the end of the contest, the payoff-relevant states are the final states $s \in \{T\} \times \mathcal{R}_S \times \mathcal{Y}$. We denote by $\pi(s)$ the payoff of a player at state s. Notice $\{\pi(s)\}_{s \in \{T\} \times \mathcal{R}_S \times \mathcal{Y}}$ can be computed (only once) outside of the model. In vector notation, we denote the vector of terminal payoffs by $\boldsymbol{\pi}$. We consider a finite grid of m values for the public scores, $Y = \{y^1, ..., y^m\}$. If a player of type θ decides to play and send a new submission, the public score of that submission is distributed according to $q_\theta(k) = \Pr(y = y^k | \theta)$, $k = 1, ..., m$. We adopt the following tie-breaking rule: The most recent draws replace older values in the top three highest public scores.

Although players are small, they have beliefs over the number of future submissions sent by their rivals. At time t, a player believes that with probability $p_t(n)$ the number of rival submissions that will arrive before the end of the competition is n. Also, the scores of those submissions are independently drawn from the distribution G, where $\Pr_G(y = y^k) = \sum_{\theta \in \Theta} \kappa(\theta) q_\theta(k)$. Furthermore, we assume that the belief about the number of rival submissions that will arrive in the future follows a Poisson distribution of parameter $\gamma(T - t)$ so

$$p_t(n) = \frac{[\gamma(T-t)]^n e^{-\gamma(T-t)}}{n!}. \tag{1}$$

Notice that under this functional form, players believe that the expected number of remaining rival's submissions, $\gamma(T - t)$, is proportional to the remaining time of the contest. To derive the expected payoff of sending an additional submission we proceed in two steps. First, we solve for the case in which a player thinks she is the last one to play, i.e., $p_t(0) = 1$, and then we solve for the the belief $p_t(n)$ given in Equation 1.

Denote by $B_t^\theta(s)$ the expected benefit of building a new submission for a player of type θ at state s, when she thinks she is the last player sending a submission before the end of the contest. For clarification, consider the following example. A player of type θ is currently at a state $s = (t, r = (1, 2), y = (y_1, y_2, y_3))$ and has an opportunity to play.

If she plays and the new submission arrives before T (which happens with probability $1 - e^{-(T-t)\lambda_\theta}$), the transition of the state depends on the score of the new submission $\tilde{y}$. The state (r, y) can transition to (r', y') where: $r' = (1, 2)$ and $y' = (y_1, y_2, y_3)$ when $\tilde{y} < y_2$; or $r' = (1, 2)$ and $y' = (y_1, \tilde{y}, y_3)$ when $y_2 \leq \tilde{y} < y_1$; or $r' = (1, 2)$ and $y' = (\tilde{y}, y_2, y_3)$ when $y_1 \leq \tilde{y}$. More generally, we can repeat these exercise for all states $s \in \mathcal{S}$ and put all these transition probabilities in a $|\mathcal{R}_S \times \mathcal{Y}| \times |\mathcal{R}_S \times \mathcal{Y}|$ matrix denoted by $\mathbf{\Omega}_\theta$. Each row of this matrix corresponds to the probability distribution over states (r', y') starting from state (r, y), conditional on the arrival of a new submission. If the new submission does not arrive, then there is no transition and the state remains (r, y). In matrix notation, where each row is a different state, the expected benefit of sending one extra submission is given by

$$\boldsymbol{B}_t^\theta = (1 - e^{-(T-t)\lambda_\theta})\mathbf{\Omega}_\theta \boldsymbol{\pi} + e^{-(T-t)\lambda_\theta}\boldsymbol{\pi}.$$

Consider a given state s. With probability $(1 - e^{-(T-t)\lambda_\theta})$ the new submission is built before the end of the contest. The score of that submission (drawn from q_θ) determines the probability distribution over final payoffs. This is given by the s-row of the matrix $\mathbf{\Omega}_\theta$. The expected payoff is computed as $(\mathbf{\Omega}_\theta)_{s\bullet} \cdot \boldsymbol{\pi}$ which corresponds to the dot-product between the probability distribution over final states starting from state s and the payoff of each terminal state. With probability $e^{-(T-t)\lambda_\theta}$ the new submission is not finished in time and therefore the final payoff for the player is given by π_s (the transition matrix is the identity matrix).[18] A player chooses to plays if and only if the expected benefit of playing net of the cost of building a submission is larger than the expected payoff of not playing, i.e.,

$$\boldsymbol{B}_t^\theta - \mathbf{c} \geq \boldsymbol{\pi} \quad \Longleftrightarrow \quad (1 - e^{-(T-t)\lambda_\theta})[\mathbf{\Omega}_\theta - \mathbf{I}]\boldsymbol{\pi} \geq \mathbf{c}. \tag{2}$$

We can now easily incorporate into Equation 2 the belief $p_t(n)$ over the number of rival submissions made after t. The final state does not depend on the order of submissions, because payoffs are realized at the end of the competition,[19] each player only cares about their ownership at the final state. Because players myopically think that they will not make another submission after the current one, we can replace the final payoff

[18]The matrix $\mathbf{\Omega}_\theta$ depends only on the probability distribution $q_\theta(\cdot)$ and can be computed outside the model.

[19]Except for ties, but we deal with this issue in the numerical implementation.

by the expected payoff after n rival's submissions and then let the agent decide to make her last submission considering this new expected payoff. That is, from state s, there is a probability distribution over $\mathcal{S}$ after n rival submissions (of scores drawn from the distribution G) given by the s-th row of matrix $\hat{\mathbf{\Omega}}^n$, where $\hat{\mathbf{\Omega}}$ is constructed similarly to $\mathbf{\Omega}$ but replacing $q_\theta(\cdot)$ by the mixture probability $g(\cdot)$. Instead of considering the payoff $\boldsymbol{\pi}$ before the last play, the player considers the expected payoff $\hat{\mathbf{\Omega}}^n\boldsymbol{\pi}$ with probability $p_t(n)$. Hence, the player plays if and only if:

$$\sum_{n=0}^{\infty}(1-e^{-(T-t)\lambda_\theta})[\mathbf{\Omega}_\theta-\mathbf{I}]\hat{\mathbf{\Omega}}^n\boldsymbol{\pi}p_t(n)\geq\mathbf{c}. \tag{3}$$

Equation 3 is similar to Equation 2, except that now the final payoff depends on how many submissions are made by rival players. Using the definition of $p_t(n)$, the definition of the exponential of a matrix,[20] and some manipulations, we obtain:

$$\Gamma_{\theta,t}\equiv(1-e^{-(T-t)\lambda_\theta})[\mathbf{\Omega}_\theta-\mathbf{I}]e^{\gamma(T-t)[\hat{\mathbf{\Omega}}-I]}\boldsymbol{\pi}\geq\mathbf{c} \tag{4}$$

Equation 4 reflects the effect of the beliefs over future rival submissions on the decision of a player to build an extra submission. Conditional on a state $s=(t,r,y)$ there are two effects driving the comparative statics on t: As the competition approaches its end, on one hand a player has less incentives to make an extra submission because it is less likely to finish building it before the end of the competition. On the other hand, it faces fewer rival submissions, which gives her more incentives to send an extra submission later on in the contest. The comparative statics on γ is intuitively clear and larger γ gives less incentives to play. The number of rival submissions that arrive on average are $\gamma(T-t)$. Therefore, the larger γ, the larger the number of submissions of the rivals and the state becomes less favorable for the player, hence the expected payoff decreases. Finally, notice that for $\theta'>\theta$ we have that $\mathbf{\Omega}_{\theta'}\succeq^{FOSD}\mathbf{\Omega}_\theta$, so better draws are more likely given the player larger incentives to play.

[20]The exponential of a matrix A is defined by $e^A\equiv\sum_{n=0}^{\infty}\frac{A^n}{n!}$

3.1 Extensions

3.1.1 Flow Cost instead of Fixed Cost

Whether players pay a fixed cost or a flow cost to build a new submission has consequences on incentives.[21] In fact, if players pay a fixed cost, then conditional on (r_i, y), sending submissions at the beginning of the contest is relatively cheaper than sending them towards the end of the contest, because new submissions are less likely to arrive closer to the end of the contest. By paying a flow cost, players may be more incentivized to play towards the end of the contest. If instead of paying a fixed cost to build new submissions players pay a flow cost while working on the new submission, Equation 4 changes to

$$(1 - e^{-(T-t)\lambda})[\mathbf{\Omega}_\theta - \mathbf{I}]e^{\gamma(T-t)[\hat{\mathbf{\Omega}}-I]}\boldsymbol{\pi} \geq \mathbf{E}[\mathbf{c}], \tag{5}$$

where

$$\text{Expected cost} \equiv \mathbf{E}[\mathbf{c}] = \int_0^{T-t} c\tau\lambda e^{-\lambda\tau}d\tau = \frac{c}{\lambda}\left[1 - e^{-\lambda(T-t)}(\lambda(T-t)+1)\right].$$

3.1.2 Forward Looking Small Players

Consider small but sequentially rational players. Each player action is either to continue or quit participating in the contest. That is, players do not have the possibility of waiting and then making submissions. They are either developing a submission or not participating in the contest at all.[22]

Given that the length of the contest is finite, the game can be solved by backward induction. At time T, no player has enough time to build a new submission. So the value of reaching state $s = (T, r, y)$ is simply $V(s) = \pi(s)$. Let V_t be a $S \times 1$ vector indicating the value at each state $s = (t, r, y)$. If the optimal decision at time t is to quit participating in the contest, then the payoff is given by

$$V_t^{\text{Quit}} = e^{\gamma(T-t)[\hat{\mathbf{\Omega}}-I]}\boldsymbol{\pi}.$$

[21] This is discussed in Loury (1979) and Lee and Wilde (1980).

[22] This strong assumption is required for identification reasons, because we cannot distinguish whether a player is working on a submission or waiting without working.

If the optimal decision is to continue playing, the expected payoff (using flow cost) is:

$$V_t^{\text{Play}} = \int_0^{T-t} \lambda e^{-\lambda\tau} \left[\mathbf{\Omega}_\theta e^{\gamma(\tau-t)[\hat{\mathbf{\Omega}}-I]} V_{t+\tau} - c\tau\right] d\tau$$

Then, we can solve by backward induction:

$$V_t = \max\left\{V_t^{\text{Quit}}, V_t^{\text{Play}}\right\}$$

Finally, one of the advantages of this simple formulation is its computational tractability. As we discuss in the next section, several elements are estimated only once, which makes fairly efficient to estimate the parameters λ and c.

4 Estimation

We estimate the privates of the model in two steps. First, we estimate a number of primitives directly from the data. Second, using the estimates of the first step, we estimate the remaining parameters using a likelihood function constructed based on the model. We repeat this procedure for each contest.

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, $\exp(\mu)$; ii) the distribution of submission arrival times, which we assume follow an exponential distribution with parameter λ, $\exp(\lambda)$; iii) the distribution of private score conditional on public score, $\{H(\cdot|p^{public})\}_{p^{public}\in[0,1]}$, 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, $Q_\theta : (0,1) \to [0,1]$, which we assume is given by the standard normal distribution, $Q_\theta(x) = \Phi\left(\frac{\log(x/(1-x))-\mu_\theta}{\sigma_\theta}\right)$; v) the distribution of types, κ, which we assume is a discrete distribution over the set of player types, Θ; vi) the time-specific distribution of the number of submissions that will be made in the reminder of the contest, $p_t(n)$, which we assume follow a Poisson distribution with parameter $\gamma(T-t)$, $p_t(n) = \frac{[\gamma(T-t)]^n e^{-\gamma(T-t)}}{n!}$; and, lastly, vii) 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), has a cumulative distribution function given by $K(c;\sigma) = c^{\sigma}$ (with $\sigma > 0$).

We estimate primitives i) through vi) in the first step, while vii) 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. We estimate iv) and v) using 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 $k = 4$ player types. Lastly, for vi), we impose that γ must equal the observed number of submissions in each contest (see Table 2), as a way of capturing that γ is an equilibrium object.

The likelihood function implied by the model is based on the decision of a player to make a new submission. Recall that a player chooses to make a new submission immediately after the arrival of each of his submissions. A player facing state variables s chooses to make a new submission at time t if and only if

$$\Gamma_{\theta,t}(s) \geq c \tag{6}$$

where $\Gamma_{\theta,t} = (1 - e^{-(T-t)\lambda_\theta})[\mathbf{\Omega}_\theta - \mathbf{I}]e^{\gamma(T-t)[\hat{\mathbf{\Omega}}-I]}\boldsymbol{\pi}$ is the vector of the net benefits of making a new submission at time t for all posible states s (before deducting the cost of making a submission) and c is the cost of a submission. $\Gamma_{\theta,t}$ depends only on primitives estimated in the first step of the estimation, which simplifies the rest of the estimation.

Based on Equation 6, we obtain that a θ-type player facing state variables s plays at time t with probability

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

Given that we do not observe the player's type, we take the expectation with respect to θ, which yields

$$\Pr(\text{play}|s, t) = \sum_{\theta} \kappa(\theta) K\left(\Gamma_{\theta,t}(s)\right),$$

where $\kappa(\theta)$ is the probability of a player being of type θ.

The likelihood is constructed using tuples $\{(s_i, t_i, t_i')\}_{i \in N}$, where i 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. 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') is given by $l(s_i, t_i, t_i') = \Pr(\text{play}|s_i, t_i) \cdot \lambda e^{(-\lambda(t_i'-t_i))}$, 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 $(s_i, t_i, t_i' > T)$ is given by $l(s_i, t_i, t_i' > T) = \Pr(\text{play}|s_i, t_i) \cdot e^{(-\lambda(T-t_i))} + 1 - \Pr(\text{play}|s_i, t_i)$, 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.

The log-likelihood function is then given by

$$L(\delta) = \sum_{i \in N} \log l(s_i, t_i, t_i'),$$

where δ is the vector of structural parameters. We perform inference using the asymptotic distribution of the maximum likelihood estimator.

4.1 Model Estimates

	μ	SE	λ	SE	σ	SE	$\log L(\hat{\delta})/N$	N
hhp	2.54	0.0691	139.3308	0.8757	0.003	0.0001	-3.5141	25316
allstate-purchase-prediction-challenge	1.9117	0.0483	86.6572	0.5533	0.0033	0.0001	-3.0259	24526
higgs-boson	2.2081	0.0523	99.8622	0.528	0.0016	0.0001	-3.2443	35772
acquire-valued-shoppers-challenge	2.0347	0.0659	123.2493	0.7765	0.0012	0.0001	-3.5276	25195
liberty-mutual-fire-peril	2.3163	0.092	84.3712	0.6932	0.0035	0.0002	-3.1449	14812
datasciencebowl	2.0146	0.0622	62.7269	0.5101	0.0043	0.0002	-2.7226	15121
axa-driver-telematics-analysis	2.0942	0.0536	98.6269	0.5193	0.0015	0.0001	-3.2953	36065
predict-west-nile-virus	2.114	0.0585	81.9751	0.4736	0.0013	0.0001	-3.1107	29965
crowdflower-search-relevance	2.0708	0.0569	68.1422	0.447	0.0024	0.0001	-2.8501	23244
caterpillar-tube-pricing	3.2151	0.0884	61.5938	0.3794	0.0024	0.0001	-2.7909	26360
liberty-mutual-group-property-inspection-prediction	2.8362	0.06	63.4536	0.2963	0.0023	0.0001	-2.8277	45875
coupon-purchase-prediction	2.1102	0.0643	66.0059	0.4856	0.0026	0.0001	-2.8118	18477
springleaf-marketing-response	2.4308	0.0515	64.4029	0.3243	0.0027	0.0001	-2.7984	39444
homesite-quote-conversion	2.2237	0.0529	81.1871	0.4257	0.003	0.0001	-3.0649	36368
prudential-life-insurance-assessment	2.1082	0.0412	72.0748	0.3379	0.0017	0.0001	-2.9028	45490
expedia-hotel-recommendations	2.2155	0.0499	40.0034	0.2655	0.0088	0.0002	-2.2063	22709

Table 6: Model Estimates (MLE), by Contest.

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

Table 6 presents the maximum likelihood estimates for the empirical model. The model was estimated separately for each contest. Column 1 shows the estimates for μ, the Poisson rate at which teams enter the competitions. The estimates suggest that the average entry time (i.e., $1/\mu$) ranges between 0.31 and 0.52 (where the contest time is normalized 1). Column 3 presents the estimates for λ, the Poisson rate at which submissions are completed. In line with Table 2, the estimates suggest that the average time between submissions ranges between 0.007 and 0.024. Column 5 presents estimates for the coefficients governing the distribution of submission costs, σ. The distribution functional form implies that the expected cost of making a submission is given by $\sigma/(1+\sigma)$. The estimates for σ suggest that the expected submission cost ranges between 0.001 and 0.008, where these submissions costs are measured relative to the total monetary rewards of each contest. Table A.4 in the Online Appendix presents the EM algorithm estimates for the type-specific distributions of scores; Table A.5 in the Online Appendix presents estimates for the distribution of private scores conditional on public scores.

With respect to how well the model fits the data, Figure 5 plots the actual number of submissions against the number of submissions that are predicted by the model. The predicted number of submissions are computed based on 2,000 simulations for 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. The contest participation predicted by the model is the average number of submissions across all of the simulations. The correlation between the actual and the predicted number of submissions is 0.72. The figure shows that the model does not systematically over or under predict participation.

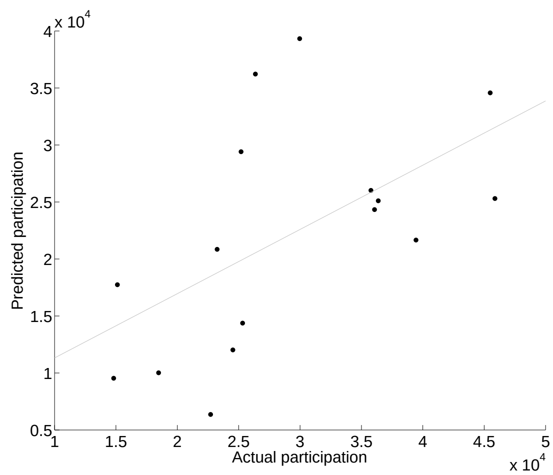


Figure 5: Number of Submissions Predicted by the Model Versus Actual Number of Submissions

Note: An observation is a contest. The coefficient of correlation between the actual and predicted number of submissions is 0.6.

5 Evaluating the Impact of Contest Design

In this section, we study how a series of counterfactual contest designs affect participation and contest outcomes. The counterfactual experiments vary the information disclosed to participants, the allocation of prizes, and impose restrictions to the number of participants.

5.1 Information Disclosure

We first study the role of information disclosure. As explained previously, the evaluation in the baseline case is based on two datasets. The contest designer chooses the size of these datasets—e.g., 60 percent of test data to generate public scores and 40 percent to generate private scores—and also which scores to disclose. In the baseline contest design, the contest designer only discloses the public scores but the final standings are computed using the private scores.

We consider two alternative designs. In one counterfactual, instead of publicly disclosing the leaderboard of public scores, the contest designer provides private feedback to each player but does not disclose a public leaderboard. That is, players observe the public score of their submissions but not the public scores of other players. In a second counterfactual, we explore the effect of eliminating the noise that generates an imperfect correlation between private and public scores. We simulate a contest where there is no noise between these scores (i.e., the public score equals the private score) and prizes are allocated according to the final standings of the public leaderboard. This second contest design can be thought of a contest where there is only one test dataset that generates both the private and public scores.

Private Feedback, no Public Leaderboard

Consider the counterfactual scenario where the contest designer does not disclose the leaderboard but players are privately informed about their public scores. Note that payoffs are realized only at the end of the contest and no information is disclosed other than the players' past submissions. Also, importantly, submissions scores are independent conditional on arriving.

Suppose that a player of type θ makes a submission. Then, starting from state s, the probability distribution over states is given by the s-th row of the matrix $\boldsymbol{\Omega_\theta}$. If after a player sends a submission, a rival player makes a submission, the probability distribution over states is given by $\hat{\boldsymbol{\Omega}}\boldsymbol{\Omega_\theta}$. Since submission scores are independent, the distribution is the same regardless of who plays first because a state is completely defined by a history of submission scores. In other words, the matrices $\boldsymbol{\Omega_\theta}$ and $\hat{\boldsymbol{\Omega}}$ commute.

Consider an information set at time t for a player of type θ, denoted by I_t^θ. Because the leaderboard is not visible, I_t^θ only contains the history of scores of this particular player. In fact, we can restrict attention to the player's largest three scores. Let $s = (t, r, y)$ be the state at time t, constructed by using only the scores of the player up to time t (contained in the history I_t), ignoring the past submissions of the rivals. From the commutativity of the transition matrices (equivalently, the stationary of distributions), the decision problem is equivalent to the decision of making a submission at time $t = 0$, starting from state s, in a contest of length $T - t$, with beliefs over the number of rivals submissions equal to $p_T(n)$. This is, we look at the row corresponding to state s in the following vector inequality:

$$(1 - e^{-(T-t)\lambda_\theta})[\mathbf{\Omega} - \mathbf{I}]e^{\gamma T[\hat{\mathbf{\Omega}} - I]}\boldsymbol{\pi} \geq \mathbf{c}$$

Incentives to make a submission change without disclosure. With the public leaderboard, each player has states where she would choose not to play and other states where she would choose to play. When not informed about the state of the contest, players cannot condition their strategy on the contest state and instead choose whether to play by computing the expected payoff of making a submission across all feasible states. A player therefore may choose to make a submission in a state where she would not make submission if she knew the state and vice versa. This possibility arises because the benefits of playing in favorable states "subsidize" the losses of playing in less favorable states. Depending on the strength of these subsidies—which depend on the player's type—a player may end up playing in a larger or smaller set of states when compared to the case when the public leaderboard is disclosed. The fact that concealing the public leaderboard may increase or decrease participation is illustrated in Example 1 below.

Figure 6 shows the impact of concealing the public leaderboard and instead only providing private feedback to the players on the number of submissions.[23] The effects are measured in percentage points and relative to the number of submissions in the baseline contest design (i.e., contest with public leaderboard). The figure shows heterogeneous effects across contests, which is not surprising given the previous discussion and the

[23]See Figure A.2 in the Online Appendix for the impact of eliminating the public leaderboard on the maximum score.

analysis in Example 1. Hiding the public leaderboard can decrease the number of submissions by as much as 40 percent (datasciencebowl) while increase it by almost 60 percent (expedia-hotel-recommendations). The figure shows that the overall number of submissions would increase by about 11 percent if the contest designer chose to hide the public leaderboard in all contests. This increase in the number of submissions translates into an increase in median maximum score.

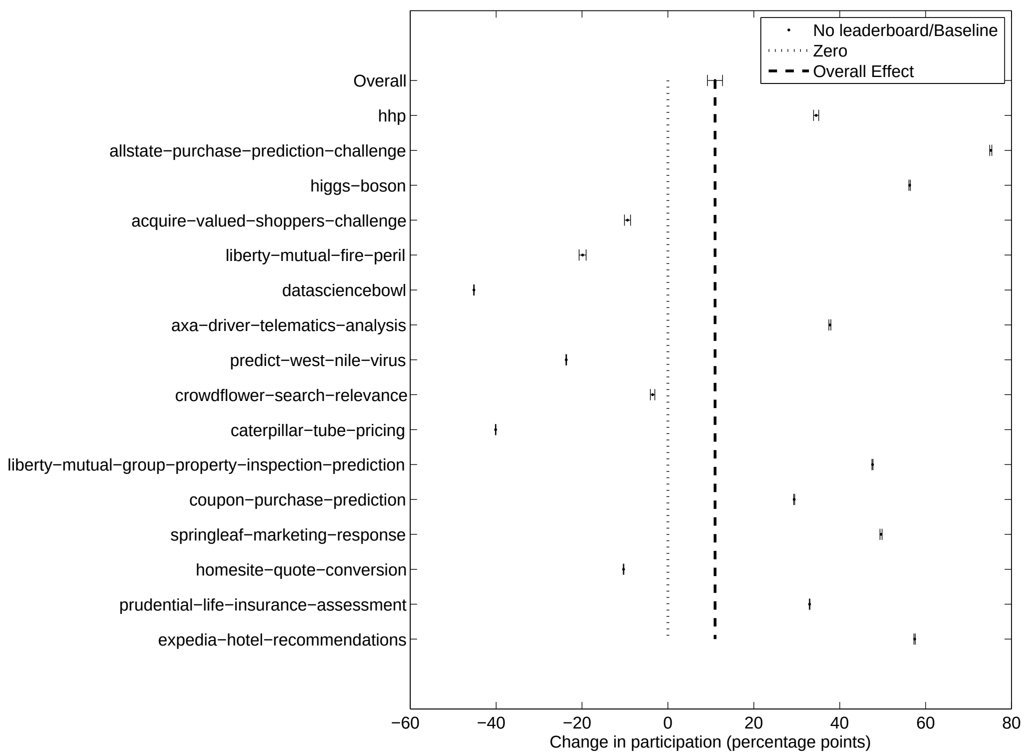
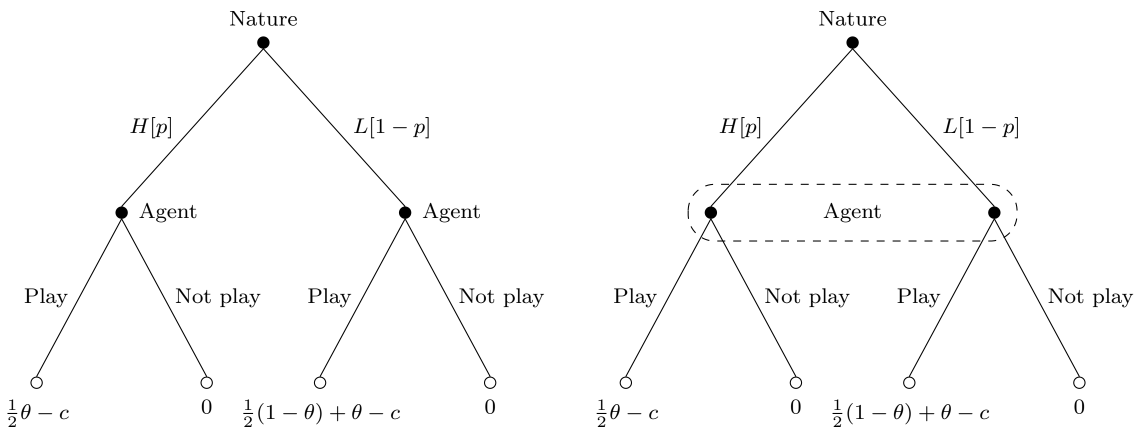


Figure 6: Change in the Number of Submissions When Comparing the Case without Leaderboard Versus the Case with Leaderboard (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

Example 1 (Disclosure of Information and Participation)**.** *Consider an agent playing against Nature. Nature draws a high score with probability* p *and a low score with probability* $1-p$*. An agent of type* $\theta \in [0,1]$ *draws a high score with probability* θ *and low score with probability* $1-\theta$*. The agent's cost of building a submission is* $c \in [0,1]$ *and we assume that ties are broken in favor of the agent with probability* $\frac{1}{2}$*. The agent* *receives $\pi = 1$ by winning the contest.*

There are two scenarios illustrated in Figure 7. A contest with a public leaderboard, and one where the draw of nature is unobserved by the agent (i.e., no leaderboard).



(a) Public Leaderboard　　**(b)** No Information Disclosure

Figure 7: Left Panel Shows the Case of a Competition with Public Leaderboard. The Right Panel Shows the Game with Private Feedback but no Leaderboard .

Consider first a contest with a public leaderboard depicted in Figure 7 (Panel A). In this case, after observing a high score the player plays if and only if $\frac{1}{2}\theta \geq c$ or $\theta > 2c$; after observing a low score the player plays if and only if $\frac{1}{2}(1-\theta) + \theta > c$ or $\theta > 2c - 1$. Consider now the a contest without a public leaderboard but with private feedback, depicted in Figure 7 (Panel B). The agent does not observe the draw of Nature but knows that high scores are drawn with probability p. The expected payoff from playing is $\frac{1}{2}p\theta + \frac{1}{2}(1-p)[\theta + 1] - c$, so the player plays if and only if $\theta \geq 2c - (1-p)$.

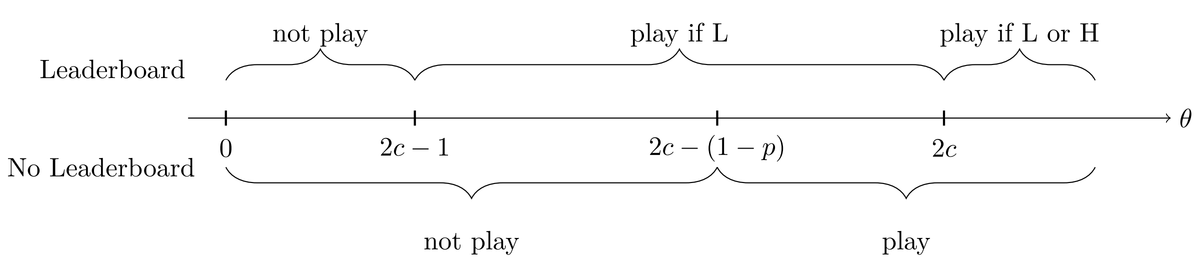


Figure 8: The effect of information disclosure on participation.

Figure 8 shows the regions where players are willing to pay to build a new submission. The figure shows that displaying the leaderboard has ambiguous effects on participation. With a public leaderboard, the player plays only if nature drew a low score and $\theta \in [2c-1, 2c-(1-p)]$*. With private feedback and no public leaderboard, however, a player of type* $\theta \in [2c-1, 2c-(1-p)]$ *would not play. Thus, for players of type* $\theta \in [2c-1, 2c-(1-p)]$ *the public leaderboard encourages participation. On the contrary, when a player is of type* $\theta \in [2c-(1-p), 2c]$ *and the public leaderboard is displayed, the player only plays if nature drew a low score. However, without a leaderboard an agent of this type would always play. Hence, displaying the leaderboard encourages participation for players of type* $\theta \in [2c-1, 2c-(1-p)]$*.*

This example shows that displaying the leaderboard may have an ambiguous effect on overall participation: it depends on how likely is an agent to draw high scores (θ), the cost of drawing, and how likely are rivals draw high scores (p). The expected participation with a public leaderboard is $(1-p)[F(2c)-F(2c-1)]+1-F(2c)$*. Without a public leaderboard the expected participation is* $1-F(2c-(1-p))$*. Hence, there is more participation on average with a public leaderboard if and only if*

$$F(p2c+(1-p)(2c-1)) \geq pF(2c)+(1-p)F(2c-1),$$

which holds if the distribution of types, F, is concave. $\square$

Perfect Correlation between Public and Private Scores

A second counterfactual design that manipulates the amount of information disclosed to participants is the contest design that eliminates the noise that causes the correlation between public and private scores to be imperfect. Eliminating this noise would require the contest designer to use 100 percent of the test data to compute the public scores, and use the public scores to determine the contest winner.

The imperfect correlation between scores distorts participation incentives as players are never certain about how well they are performing in the contest. On the one hand, the imperfect correlation encourages participants because all players have a chance at winning the contest, but on the other hand, it discourages players at the very top because leading the public leaderboard is not sufficient to win the contest.[24]

[24]It is also worth mentioning that computing the public score with 100 percent of the test data has

Figure 9 shows the impact of eliminating the noise between the public and private scores (i.e., public score equals private score) on the number of submissions. The effects are measured in percentage points and relative to the number of submissions in the baseline contest design (i.e., contest with public leaderboard and noise). The figure shows that the number of submissions decreases by about 2.7 percent on average when eliminating the noise but with heterogeneous effects. This decrease in the number of submissions translates into a decrease in the average maximum score (see Figure A.3 in the Online Appendix).

Noteworthy is the fact that both counterfactual designs that we consider in this subsection provide consistent results. Hiding the public leaderboard can be thought of as an extremely noisy public leaderboard, which is the opposite case of allocating prizes according to the final standings of the public leaderboard (an extremely informative public leaderboard). In both comparisons, we find that more noise on average increases participation in economically significant magnitudes.

---

the potential drawback that participants may attempt to engage in *overfitting* and submit solutions that maximize the public score but are not robust outside of the test data. We abstract away from this potential effect.

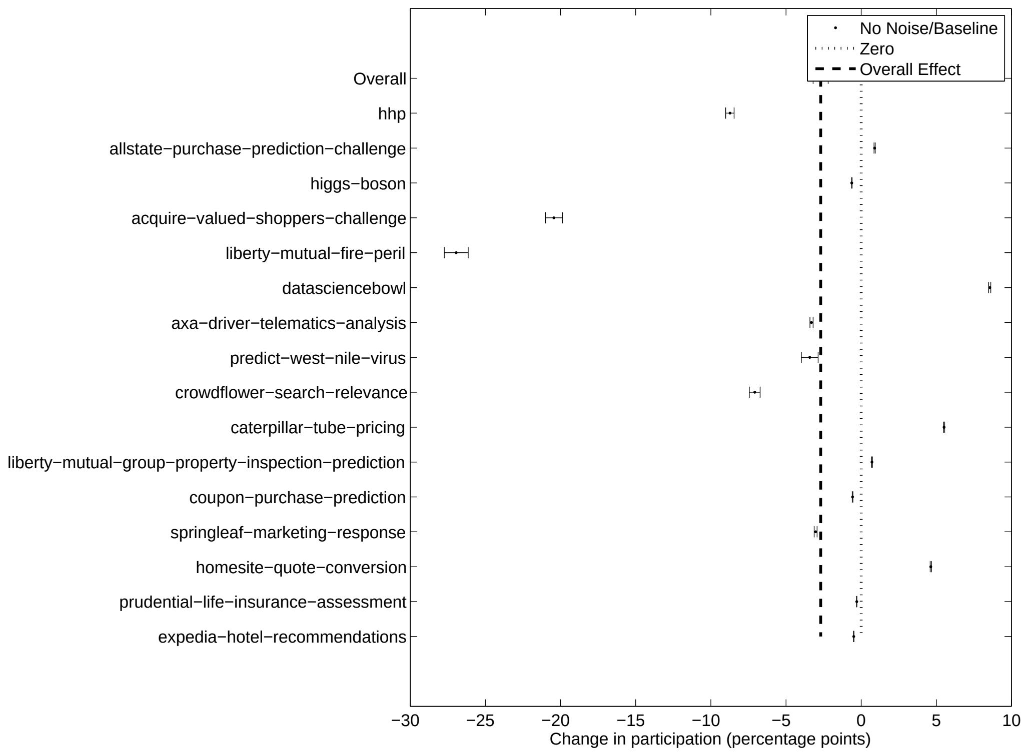


Figure 9: Change in the Number of Submissions When Comparing the Case Without Leaderboard Noise Versus the Case with Leaderboard Noise (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

5.2 Number of Prizes

We next analyze the role of the allocation of prizes on contest outcomes. Instead of allocating prizes to the j-highest ranked players, we simulate a contest that allocates a single prize to the winner, keeping the total award money fixed. The optimal allocation of prizes has been explored in several articles, including Lazear and Rosen (1979), Taylor (1995), Moldovanu and Sela (2001), Che and Gale (2003), Sisak (2009), Olszewski and Siegel (2015), Kireyev (2016), Xiao (2016), Strack (2016), and Balafoutas et al. (2017). The literature has found that the shape of the cost function plays an important role in determining the optimal prize allocation when the contest designer's goal is to maximize aggregate effort.

Figure 10 shows the impact of awarding a single prize on the number of submissions.[25] The effects are measured in percentage points and relative to the number of submissions in the baseline contest design (i.e., contest with public leaderboard and three prizes). Our results show that changing the allocation of prizes has a small and statistically insignificant overall effect on participation. While the results are heterogeneous across contests, the magnitude of the effect is less than 1 percent (in absolute value) in all but one contest. To explain this result, notice that the first order effect on incentives is whether or not a player is ranked among the first three players at the end of the contest. Conditional on that event, the effect of allocating one or three prizes is small because of the uncertainty created by the imperfect correlation between public and private scores.

[25] Figure A.4 in the Online Appendix displays the impact of offering a single prize on the maximum score.

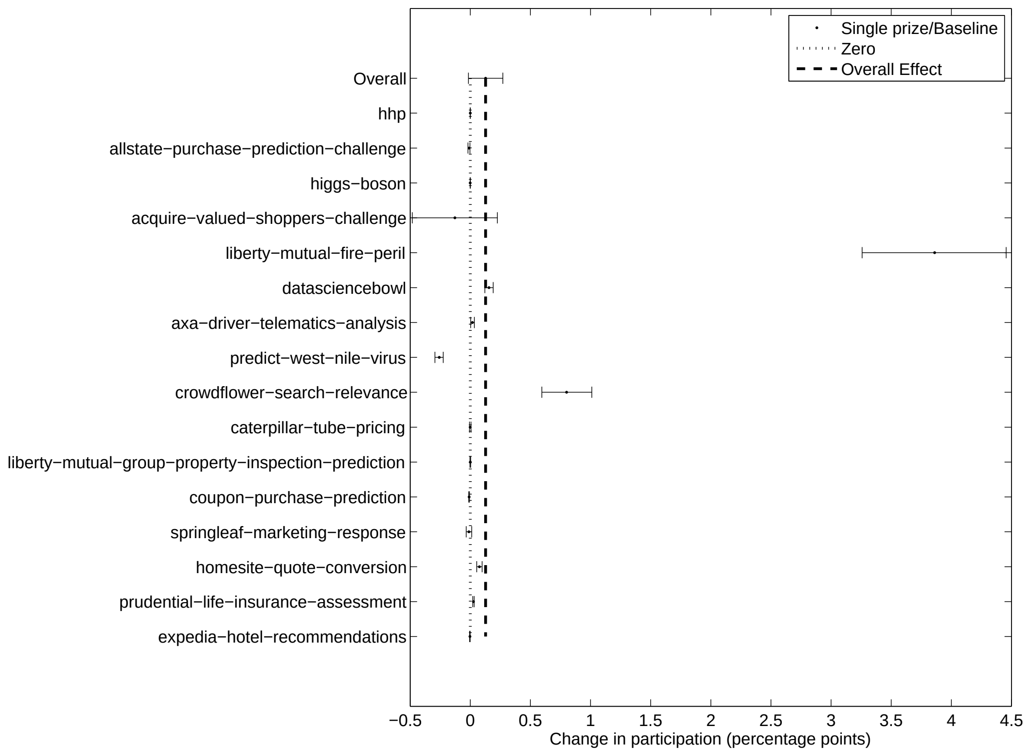


Figure 10: Change in the Number of Submissions When Comparing the Case with One Prize Versus the Case with Three Prizes (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

5.3 Limiting the Number of Participants

Lastly, we study the role of limiting participation on contest outcomes. The literature has isolated two effects caused by limiting participation. On the one hand, limiting participation reduces the number of participants exerting effort (in our case, sending submissions). On the other hand, with fewer players, the marginal benefit of effort increases as players face less competition. Whether the effect on the effort of players compensates for the reduction in the number of players is ex-ante ambiguous. Some articles have argued that limiting participation may be optimal (see Che and Gale 2003; Kireyev 2016; Taylor 1995) although these results are generally based on models

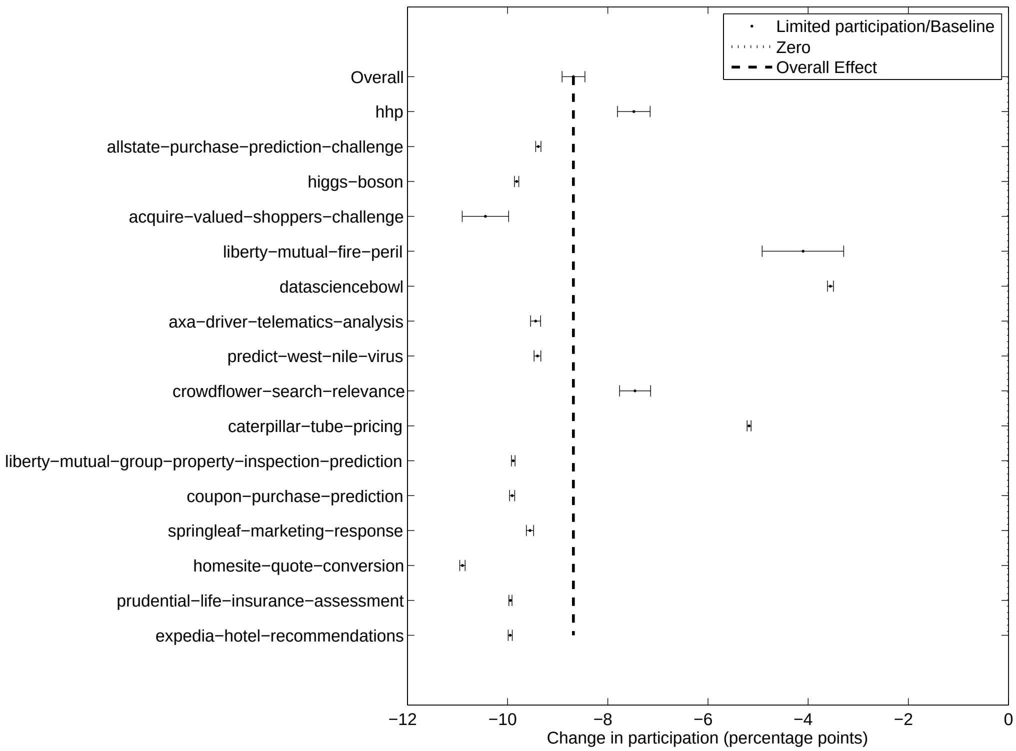


Figure 11: Change in the Number of Submissions When Comparing the Case of a Restricted Participation Contest with 90 Percent of the Teams Versus the Case with the Observed Number of Teams (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

of static contests. We contribute to the literature by shedding light on the impact of limiting participation in a dynamic environment with heterogeneous players.

We first consider the case where we reduce the number of participants in each contest by 10 percent, keeping the distribution of player types unchanged. Figure 11 shows how participation changes for each contest with such intervention.[26] The effects are measured in percentage points and relative to the number of submissions in the baseline contest design (i.e., contest with public leaderboard and full participation). We find that the increase in individual incentives caused by reduced competition does not fully compensate for the reduction in the number of players, leading to an average decrease in the total number of submissions of 8.7 percent.

We also consider the case when we decrease the number of participants in each contest by 10 percent but change the distribution of player types to be composed by "high types" or experts only.[27] This exercise is motivated by the fact that Kaggle in some instances does restrict competitions to more experienced players. We present the results in Figure A.6 in the Online Appendix and again find that reducing the number of participants leads to an average decrease in the number of submissions. In this case, however, the average decrease in the number of submissions exceeds 10 percent because of how the change in the composition of players discourages participants.

5.4 Robustness

With respect to the robustness of the results presented in this section, we study how the contest design effects that we estimate change when perturbing the model. In particular, we consider how the results change when using the specification where players pay a flow cost rather than a fixed cost when building a submission (see the discussion in Section 3.1).[28] It has been argued in the literature that whether players pay a flow or fixed cost

[26] Figure A.5 in the Online Appendix displays the impact of limited participation on the maximum score.

[27] We define "high type" as the player type with the greatest $\mu + 2\sigma$ value (see the definition of Q_θ in Section 4).

[28] The estimates of the flow cost specification of the model are presented in Table A.6 in the Online Appendix. The fit of the model is slightly worse when using the flow cost specification.

matters for players' incentives (see for instance Loury 1979 and Lee and Wilde 1980). Table A.7 in the Online Appendix shows the results of each of the counterfactuals in this section for both the fixed cost and flow cost specifications. The results are qualitatively identical throughout.

6 Discussion

We study dynamic prediction contests and explore the effect of contest design on participation incentives and contest outcomes. We provide two main contributions.

Our first contribution is to introduce a novel and tractable empirical model—i.e., a model with a state space that is computationally manageable—to study a setting where players can submit multiple submissions throughout the contest. The contest designer displays, in real time, a public leaderboard in which participants observe a noisy signal of their position and use it to decide whether to continue participating or to quit the contest. Our model relies on various simplifications, which are motivated by the empirical evidence, but is general enough to be applied to other settings. The computational tractability is achieved by assuming that players are small, i.e., they do not consider the effect of their actions on rival player's strategies. However, players form beliefs, which are correct in equilibrium, about the future number of submissions in the contest. In our framework, the expected payoff for set of beliefs about the number of rivals' submissions can be computed efficiently by computing the exponential of a matrix. It is this tractability that allow us to easily estimate the parameters of the model and to compute outcomes under counterfactual contest designs.

Our second contribution is to shed light on contest design in a dynamic setting with heterogeneous players. We simulate several counterfactual scenarios to explore alternative contest designs. In particular, we study the role of information disclosure, allocation of prizes, and limited participation in shaping the player's incentives. Although we find heterogeneity in the results, changes in information disclosure induce the greatest participation response. Specifically, we find that only providing private feedback instead of publicly disclosing the leaderboard would increase the average number of submissions in a contest by 11 percent.

7 References

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

Bajari, Patrick, C Lanier Benkard, and Jonathan Levin (2007) "Estimating dynamic models of imperfect competition," *Econometrica*, Vol. 75, pp. 1331–1370.

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.

Baker, George P, Michael C Jensen, and Kevin J Murphy (1988) "Compensation and incentives: Practice vs. theory," *The journal of Finance*, Vol. 43, pp. 593–616.

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.

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

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

Bockstedt, Jesse, Cheryl Druehl, and Anant Mishra (2016) "Heterogeneous Submission Behavior and its Implications for Success in Innovation Contests with Public Submissions," *Production and Operations Management*.

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.

Chawla, Shuchi, Jason D Hartline, and Balasubramanian Sivan (2015) “Optimal crowdsourcing contests,” *Games and Economic Behavior*.

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

Chen, Min, Shiwen Mao, and Yunhao Liu (2014) “Big data: a survey,” *Mobile Networks and Applications*, Vol. 19, pp. 171–209.

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

Corchón, Luis C (2007) “The theory of contests: a survey,” *Review of Economic Design*, Vol. 11, pp. 69–100.

Diamond, Peter A (1971) “A model of price adjustment,” *Journal of economic theory*, Vol. 3, pp. 156–168.

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) “Auctionin entry into tournaments,” *Journal of Political Economy*, Vol. 107, pp. 573–605.

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

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

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

Hu, Zhenghui and Wenjun Wu (2015) “Game Theoretic Analysis for Offense-Defense Challenges of Algorithm Contests on TopCoder,” in *Service-Oriented System Engineering (SOSE), 2015 IEEE Symposium on*, pp. 339–346, IEEE.

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.

Jofre-Bonet, Mireia and Martin Pesendorfer (2003) "Estimation of a dynamic auction game," *Econometrica*, Vol. 71, pp. 1443–1489.

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

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

Konrad, Kai A (2012) "Dynamic contests and the discouragement effect," *Revue d'économie politique*, Vol. 122, pp. 233–256.

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.

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

Lee, Tom and Louis L Wilde (1980) "Market structure and innovation: a reformulation," *The Quarterly Journal of Economics*, Vol. 94, pp. 429–436.

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

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

Loury, Glenn C (1979) "Market structure and innovation," *The quarterly journal of economics*, pp. 395–410.

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."

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 us 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.

Wright, Brian D (1983) "The economics of invention incentives: Patents, prizes, and research contracts," *The American Economic Review*, Vol. 73, pp. 691–707.

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

Yang, Yang, Pei-yu Chen, and Paul Pavlou (2009) "Open innovation: An empirical study of online contests," *ICIS 2009 Proceedings*, p. 13.

Online Appendix: Not For Publication

Dynamic Tournament Design: An Application to Prediction Contests

Jorge Lemus and Guillermo Marshall

Name of the Competition	Total Reward	Number of Submissions	Teams	Start Date	Deadline
Predict Grant Applications	5,000	2,800	204	12/13/2010	02/20/2011
RTA Freeway Travel Time Prediction	10,000	3,129	356	11/23/2010	02/13/2011
Deloitte/FIDE Chess Rating Challenge	10,000	1,563	181	02/07/2011	05/04/2011
Heritage Health Prize	500,000	25,316	1,353	04/04/2011	04/04/2013
Wikipedia's Participation Challenge	10,000	1,020	90	06/28/2011	09/20/2011
Allstate Claim Prediction Challenge	10,000	1,278	102	07/13/2011	10/12/2011
dunnhumby's Shopper Challenge	10,000	1,872	277	07/29/2011	09/30/2011
Give Me Some Credit	5,000	7,730	925	09/19/2011	12/15/2011
Don't Get Kicked!	10,000	7,261	570	09/30/2011	01/05/2012
Algorithmic Trading Challenge	10,000	1,406	111	11/11/2011	01/08/2012
What Do You Know?	5,000	1,747	239	11/18/2011	02/29/2012
Photo Quality Prediction	5,000	1,356	200	10/29/2011	11/20/2011
KDD Cup 2012, Track 1	8,000	13,076	657	02/20/2012	06/01/2012
KDD Cup 2012, Track 2	8,000	5,276	163	02/20/2012	06/01/2012
Predicting a Biological Response	20,000	8,837	699	03/16/2012	06/15/2012
Online Product Sales	22,500	3,755	363	05/04/2012	07/03/2012
EMI Music Data Science Hackathon - July 21st - 24 hours	10,000	1,319	133	07/21/2012	07/22/2012
Belkin Energy Disaggregation Competition	25,000	1,526	165	07/02/2013	10/30/2013
Merck Molecular Activity Challenge	40,000	2,979	236	08/16/2012	10/16/2012
U.S. Census Return Rate Challenge	25,000	2,666	243	08/31/2012	11/11/2012
Amazon.com - Employee Access Challenge	5,000	16,872	1,687	05/29/2013	07/31/2013
The Marinexplore and Cornell University Whale Detection Challenge	10,000	3,293	245	02/08/2013	04/08/2013
See Click Predict Fix - Hackathon	1,000	1,051	80	09/28/2013	09/29/2013
KDD Cup 2013 - Author Disambiguation Challenge (Track 2)	7,500	2,304	237	04/19/2013	06/12/2013
Influencers in Social Networks	2,350	2,105	132	04/13/2013	04/14/2013
Personalize Expedia Hotel Searches - ICDM 2013	25,000	3,502	337	09/03/2013	11/04/2013
StumbleUpon Evergreen Classification Challenge	5,000	7,495	625	08/16/2013	10/31/2013
Personalized Web Search Challenge	9,000	3,570	194	10/11/2013	01/10/2014
See Click Predict Fix	4,000	5,570	532	09/29/2013	11/27/2013
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	25,195	952	04/10/2014	07/14/2014
The Hunt for Prohibited Content	25,000	4,992	285	06/24/2014	08/31/2014
Liberty Mutual Group - Fire Peril Loss Cost	25,000	14,812	634	07/08/2014	09/02/2014
Tradeshift Text Classification	5,000	5,648	375	10/02/2014	11/10/2014
Driver Telematics Analysis	30,000	36,065	1,528	12/15/2014	03/16/2015
Diabetic Retinopathy Detection	100,000	7,002	661	02/17/2015	07/27/2015
Click-Through Rate Prediction	15,000	31,015	1,604	11/18/2014	02/09/2015
Otto Group Product Classification Challenge	10,000	43,525	3,514	03/17/2015	05/18/2015
Crowdflower Search Results Relevance	20,000	23,244	1,326	05/11/2015	07/06/2015
Avito Context Ad Clicks	20,000	5,949	414	06/02/2015	07/28/2015
ICDM 2015: Drawbridge Cross-Device Connections	10,000	2,355	340	06/01/2015	08/24/2015
Caterpillar Tube Pricing	30,000	26,360	1,323	06/29/2015	08/31/2015
Liberty Mutual Group: Property Inspection Prediction	25,000	45,875	2,236	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	39,444	2,226	08/14/2015	10/19/2015
Truly Native?	10,000	3,223	274	08/06/2015	10/14/2015
Homesite Quote Conversion	20,000	36,368	1,764	11/09/2015	02/08/2016
Prudential Life Insurance Assessment	30,000	45,490	2,619	11/23/2015	02/15/2016
BNP Paribas Cardif Claims Management	30,000	54,516	2,926	02/03/2016	04/18/2016
Home Depot Product Search Relevance	40,000	35,619	2,125	01/18/2016	04/25/2016
Santander Customer Satisfaction	60,000	93,559	5,123	03/02/2016	05/02/2016
Expedia Hotel Recommendations	25,000	22,709	1,974	04/15/2016	06/10/2016
Avito Duplicate Ads Detection	20,000	8,153	548	05/06/2016	07/11/2016
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.

Public Ranking of Winner	Frequency	Probability	Cumulative Probability
1	27	49.09	49.09
2	12	21.82	70.91
3	3	5.45	76.36
4	6	10.91	87.27
5	1	1.82	89.09
6	2	3.64	92.73
11	3	5.45	98.18
54	1	1.82	100.00

Table A.2: Public Leaderboard Ranking of Competition Winners

Number of Competitions	Frequency	Probability	Cumulative Probability
1	23,443	71.75	71.75
2	4,577	14.01	85.76
3	1,861	5.70	91.46
4	903	2.76	94.22
5 or more	1,887	5.78	100.00

Table A.3: Number of Competitions by User

	Type 1			Type 2			Type 3			Type 4				
	μ_1	σ_1	κ_1	μ_2	σ_2	κ_2	μ_3	σ_3	κ_3	μ_4	σ_4	κ_4	$\log L(\hat{\delta})/N$	N
hhp	4.0424	0.6251	0.5965	4.5966	0.3813	0.2479	3.1150	0.4779	0.0402	2.8849	0.4540	0.1153	-0.9037	26646
allstate-purchase-prediction-challenge	4.4368	0.1139	0.0000	3.8591	1.3087	0.5552	4.2364	0.0000	0.1090	1.4724	2.0368	0.3358	1.2919	27046
higgs-boson	2.0874	1.0490	0.5864	-0.6665	0.6531	0.0636	0.1206	0.9316	0.3315	2.2643	0.3671	0.0184	-1.6298	38098
acquire-valued-shoppers-challenge	0.9063	0.6676	0.6857	-0.0689	0.0000	0.0758	-0.0659	0.0362	0.0371	1.6983	0.6098	0.2014	-0.5635	26870
liberty-mutual-fire-peril	1.0140	0.6202	0.2404	2.3703	0.5779	0.0973	0.8659	0.6470	0.0724	1.9080	0.7278	0.5899	-1.2011	15443
datasciencebowl	1.8207	0.2902	0.1798	2.0250	0.4657	0.1004	2.7798	0.6750	0.5046	3.5648	0.4480	0.2152	-1.1421	16579
axa-driver-telematics-analysis	-0.6523	0.0001	0.0015	0.9715	1.0128	0.7334	-0.6522	0.0000	0.0969	-0.6096	0.0783	0.1682	-0.8409	38694
predict-west-nile-virus	1.0158	0.4702	0.4544	1.2849	0.1129	0.0115	0.2743	0.0580	0.0602	1.7279	0.5264	0.4739	-0.9371	32179
crowdflower-search-relevance	1.1752	0.6986	0.2858	2.0954	0.5608	0.0618	1.0083	0.6603	0.0751	2.2824	0.7492	0.5773	-1.2721	24633
caterpillar-tube-pricing	4.9893	0.3346	0.0814	4.6638	0.1080	0.0015	7.4526	0.8554	0.5396	6.2479	0.9127	0.3775	-1.4185	28622
liberty-mutual-group-property-inspection-prediction	1.4246	0.3908	0.0300	2.9764	0.8699	0.3082	4.1295	0.8965	0.4781	2.1768	0.6682	0.1837	-1.5103	48403
coupon-purchase-prediction	0.1576	0.2326	0.0068	0.2612	0.2552	0.0659	0.3295	0.1530	0.0076	-0.6254	1.6139	0.9198	-1.7620	19446
springleaf-marketing-response	1.6692	0.4889	0.0800	2.9008	0.4245	0.1326	3.8894	0.6918	0.3755	2.8837	0.7086	0.4120	-1.2750	42780
homesite-quote-conversion	5.0174	1.0821	0.8271	5.6153	0.1494	0.0521	2.4063	0.7822	0.0669	1.5355	0.2046	0.0539	-1.3467	38284
prudential-life-insurance-assessment	1.5646	0.3658	0.0727	3.8653	0.9771	0.6364	1.2810	0.3357	0.0424	2.1279	0.5712	0.2484	-1.5459	49491
expedia-hotel-recommendations	1.5657	0.0464	0.0512	0.2111	1.2160	0.5392	1.5653	0.0265	0.0379	1.3177	0.5655	0.3717	-0.7960	24348

Table A.4: EM Algorithm Estimates for the Type-specific Distribution of Scores, q_θ.

Note: The model is estimated separately for each contest. μ_i and σ_i are the parameters in type i's distribution of scores $Q_i(s) = \Phi\left(\frac{\log(s/(1-s))-\mu_i}{\sigma_i}\right)$. κ_i is the fraction of players of type i. $\log L(\hat{\delta})/N$ is the value of the log-likelihood function evaluated at the EM estimates. Standard errors are available.

	α	SE	β	SE	N
hhp	0.0022	0.0717	1.0025	0.0737	25316
allstate-purchase-prediction-challenge	0.0005	0.0197	1.0043	0.0221	24526
higgs-boson	-0.0002	0.0183	1.0168	0.0224	35772
acquire-valued-shoppers-challenge	-0.0139	0.033	1.0105	0.043	25195
liberty-mutual-fire-peril	0.044	0.0449	0.9083	0.054	14812
datasciencebowl	0.0011	0.0562	0.9987	0.0606	15121
axa-driver-telematics-analysis	-0.0019	0.0179	1.0019	0.0253	36065
predict-west-nile-virus	0.0237	0.0444	0.9756	0.0548	29965
crowdflower-search-relevance	0.0174	0.0362	0.986	0.0426	23244
caterpillar-tube-pricing	-0.014	0.3918	1.014	0.3929	26360
liberty-mutual-group-property-inspection-prediction	0.0061	0.0383	0.9961	0.0407	45875
coupon-purchase-prediction	0.033	0.0134	0.9022	0.0279	18477
springleaf-marketing-response	0.0092	0.052	0.9894	0.0553	39444
homesite-quote-conversion	0.0028	0.0401	0.997	0.0417	36368
prudential-life-insurance-assessment	0.0092	0.042	0.9933	0.0447	45490
expedia-hotel-recommendations	0.0006	0.019	0.9983	0.0269	22709

Table A.5: Maximum Likelihood Estimates of the Distribution of Private Scores Conditional on Public Scores, by Contest. 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.

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

	μ	SE	λ	SE	σ	SE	$\log L(\hat{\delta})/N$	N
hhp	2.54	0.0691	139.3308	0.8757	0.0027	0.0001	-3.4837	25316
allstate-purchase-prediction-challenge	1.9117	0.0483	86.6572	0.5533	0.0028	0.0001	-3.0087	24526
higgs-boson	2.2081	0.0523	99.8622	0.528	0.0017	0.0001	-3.234	35772
acquire-valued-shoppers-challenge	2.0347	0.0659	123.2493	0.7765	0.0014	0.0001	-3.5241	25195
liberty-mutual-fire-peril	2.3163	0.092	84.3712	0.6932	0.0024	0.0002	-3.0775	14812
datasciencebowl	2.0146	0.0622	62.7269	0.5101	0.0028	0.0002	-2.693	15121
axa-driver-telematics-analysis	2.0942	0.0536	98.6269	0.5193	0.0015	0.0001	-3.2928	36065
predict-west-nile-virus	2.114	0.0585	81.9751	0.4736	0.001	0.0001	-3.1037	29965
crowdflower-search-relevance	2.0708	0.0569	68.1422	0.447	0.0023	0.0001	-2.7544	23244
caterpillar-tube-pricing	3.2151	0.0884	61.5938	0.3794	0.0019	0.0001	-2.7789	26360
liberty-mutual-group-property-inspection-prediction	2.8362	0.06	63.4536	0.2963	0.0018	0.0001	-2.8166	45875
coupon-purchase-prediction	2.1102	0.0643	66.0059	0.4856	0.0025	0.0001	-2.7249	18477
springleaf-marketing-response	2.4308	0.0515	64.4029	0.3243	0.0023	0.0001	-2.7805	39444
homesite-quote-conversion	2.2237	0.0529	81.1871	0.4257	0.0021	0.0001	-3.0516	36368
prudential-life-insurance-assessment	2.1082	0.0412	72.0748	0.3379	0.0019	0.0001	-2.9009	45490
expedia-hotel-recommendations	2.2155	0.0499	40.0034	0.2655	0.0049	0.0003	-1.9312	22709

Table A.6: Maximum Likelihood Estimates for the Flow Cost Specification of the Model, by Contest.

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

Case	Fixed cost	Flow cost
No leaderboard	11.00	12.83
	(9.23,12.74)	(10.95,14.75)
Single prize	0.13	-0.06
	(-0.01,0.27)	(-0.19,0.07)
No noise	-2.69	-2.60
	(-3.20,-2.19)	(-3.10,-2.11)
Limited participation	-8.69	-8.80
	(-8.91,-8.46)	(-9.01,-8.58)
Experts	-12.66	-12.40
	(-13.37,-11.95)	(-13.12,-11.67)

Table A.7: Comparing the Effects of Contest Design on Participation by Model Specification. The Participation Effects of Counterfactual Designs are Measured in Percentage Points Relative to the Baseline Contest Design.

Note: The table includes estimates for all contests except for the Santander Customer Satisfaction contest. Bootstrapped 95 percent confidence intervals in parenthesis.

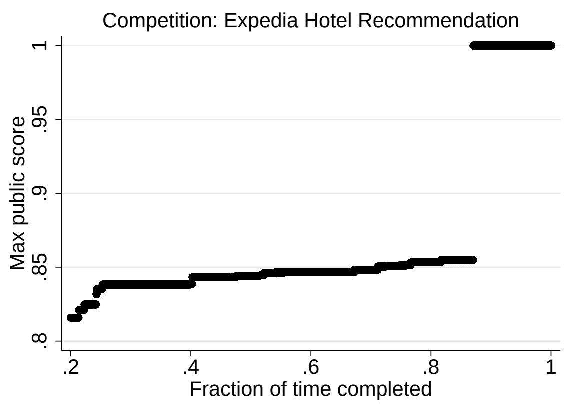


Figure A.1: Evolution of the Maximum Public Score in the Expedia Hotel Recommendation Contest. The Jump in the Maximum Public Score Captures a Drastic Submission

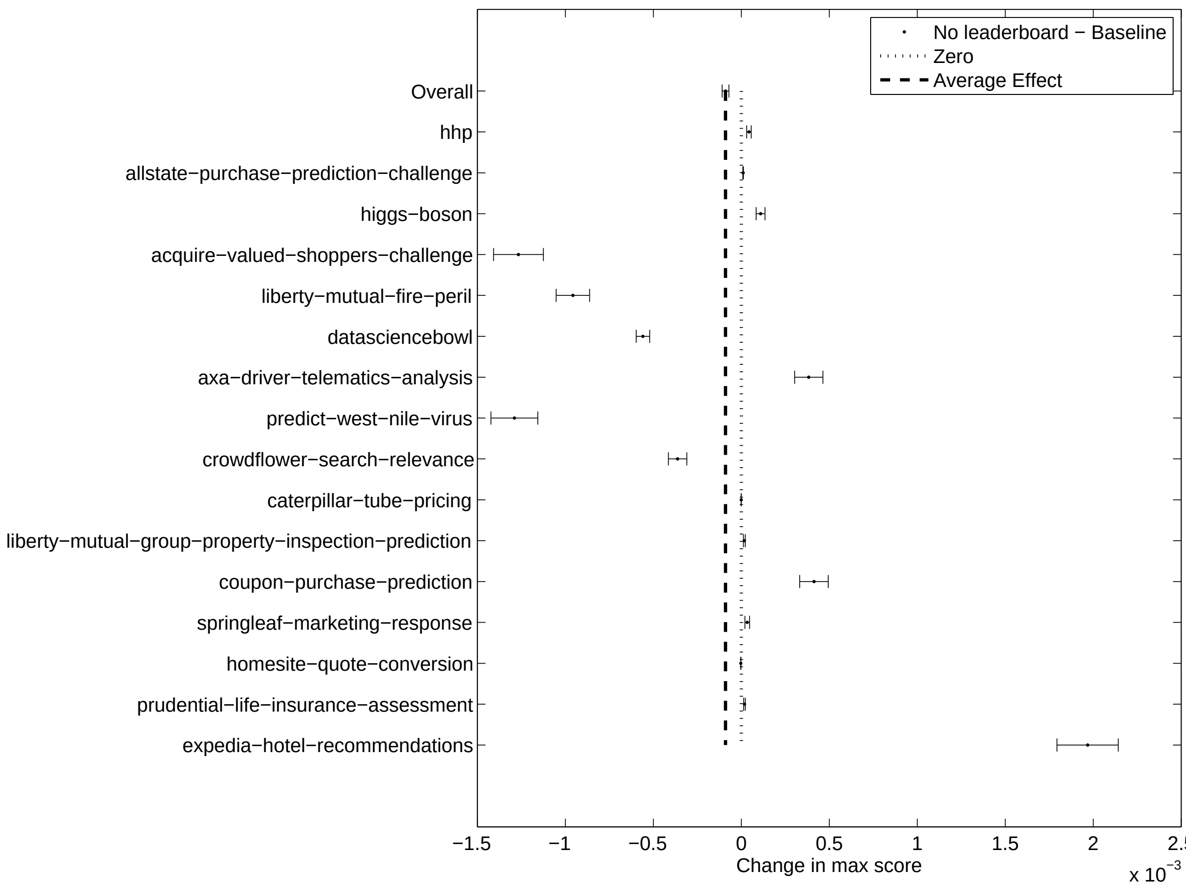


Figure A.2: Change in the Maximum Score when Comparing the Case without Leaderboard Versus the Case with Leaderboard (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

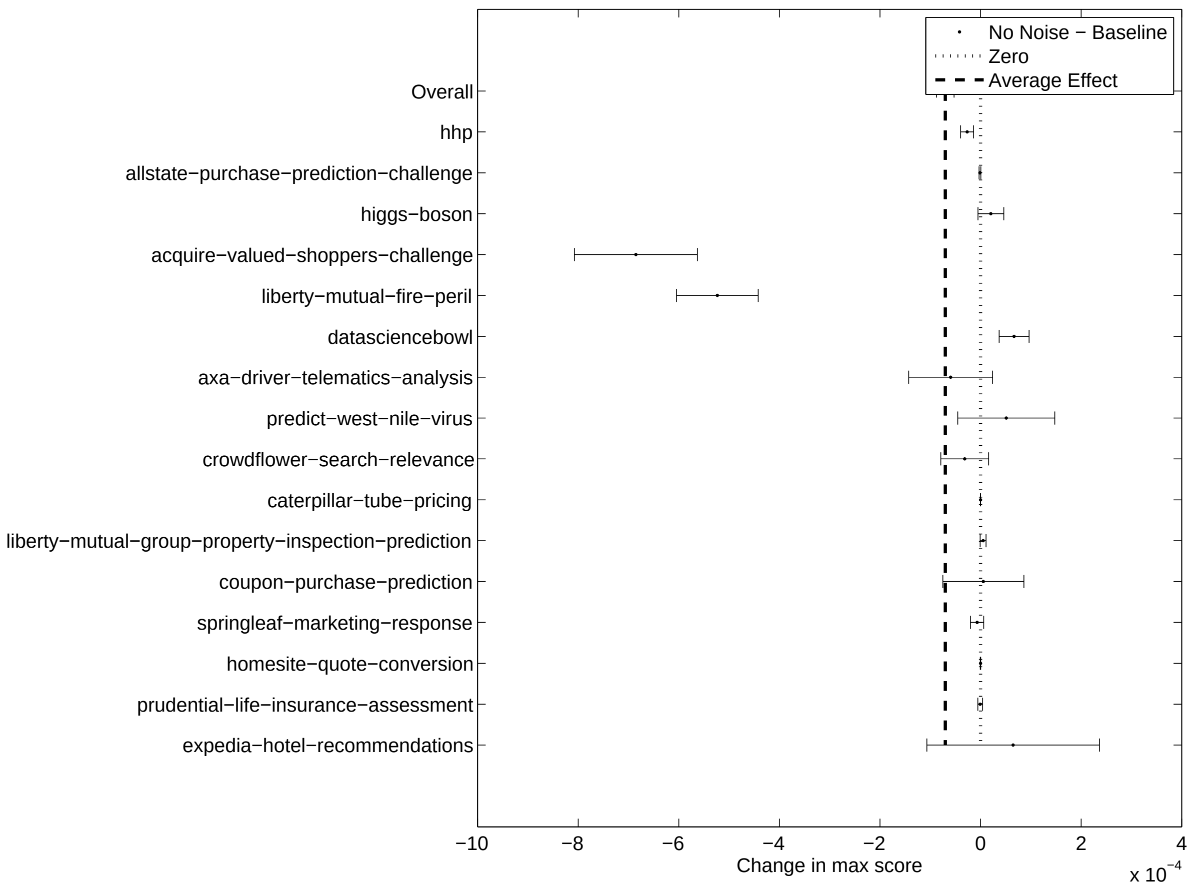


Figure A.3: Change in the Maximum Score when Comparing the Case without Leaderboard Noise Versus the Case with Leaderboard Noise (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

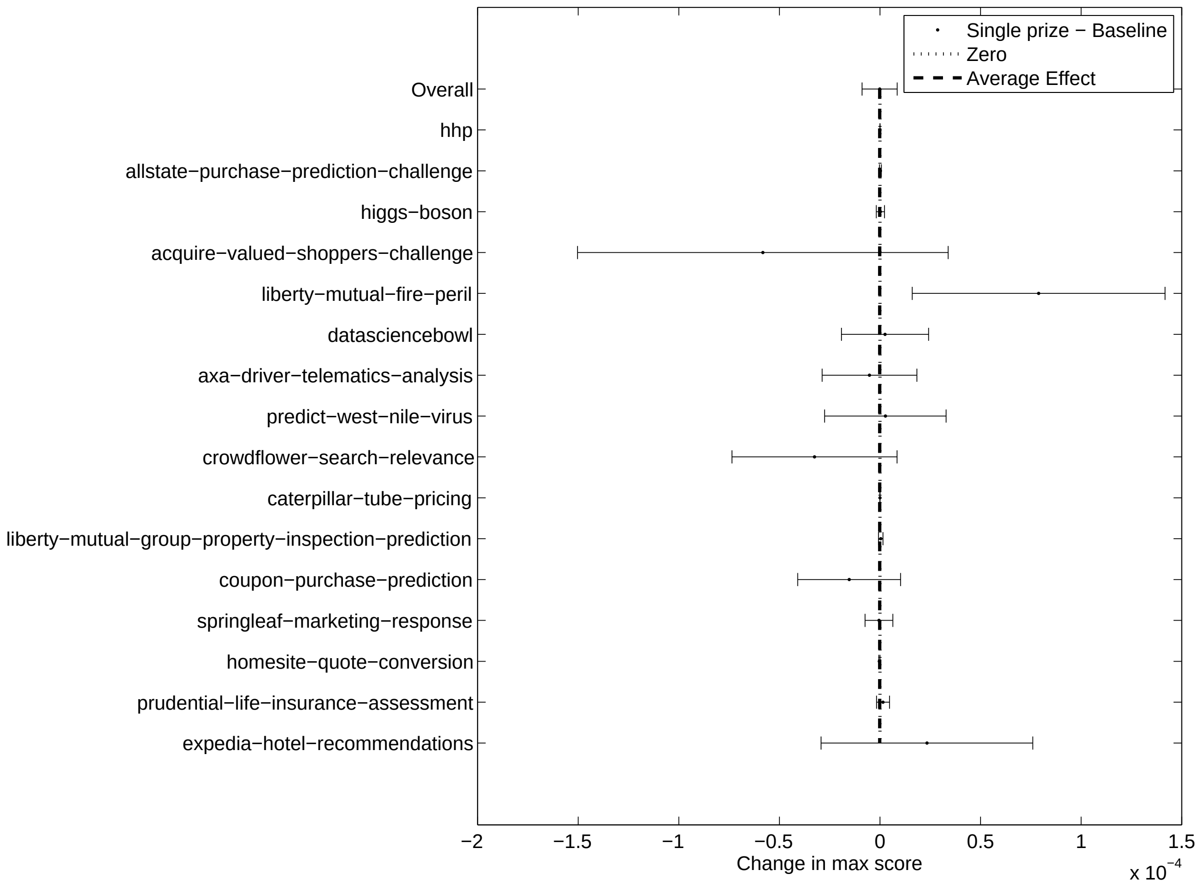


Figure A.4: Change in the Maximum Score when Comparing the Case with One Prize Versus the Case with Three Prizes (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

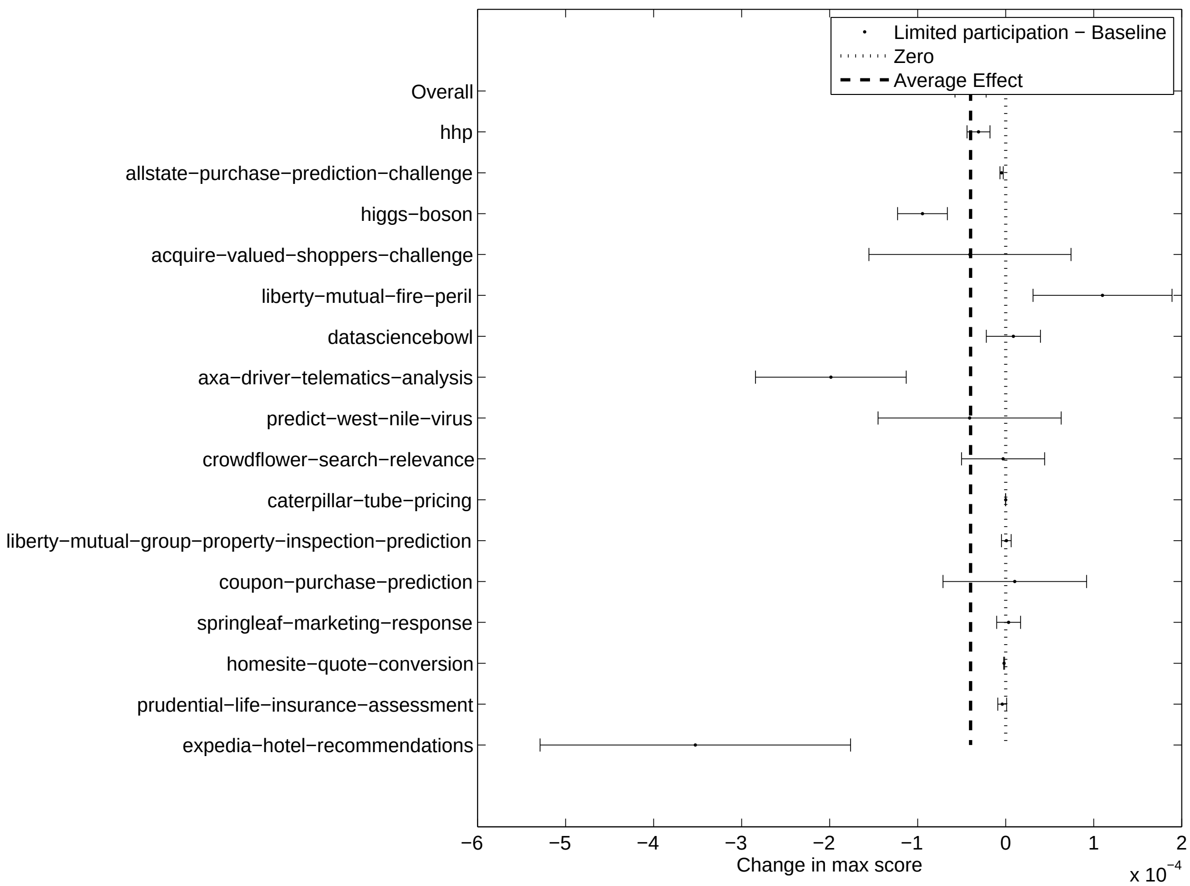


Figure A.5: Change in the Maximum Score when Comparing the Case of a Restricted Participation Contest with 90 Percent of the Teams Versus the Case with the Observed Number of Teams (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

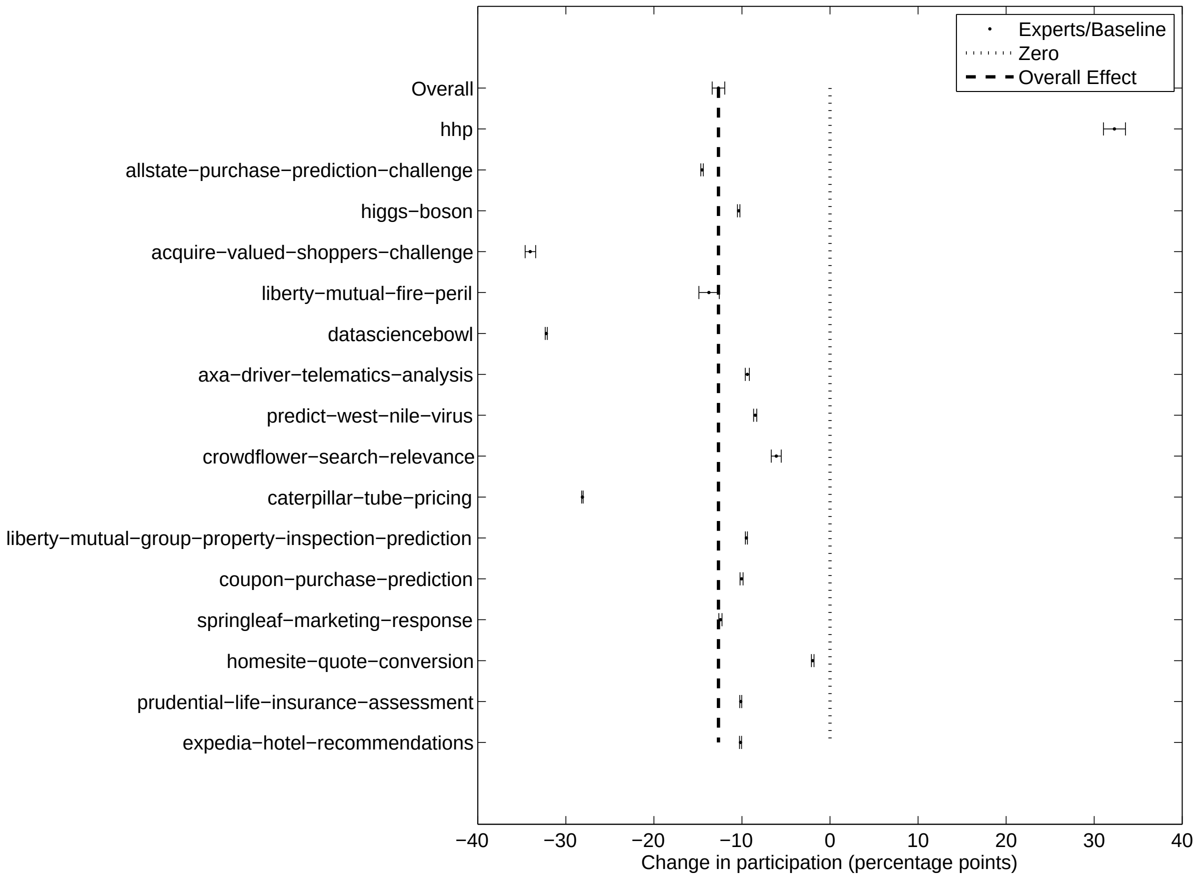


Figure A.6: Change in the Number of Submissions when Comparing the Case of a Restricted Participation Contest with 90 Percent of the Teams—All of Them of the Highest Player Type—Versus the Case with the Observed Number of Teams (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

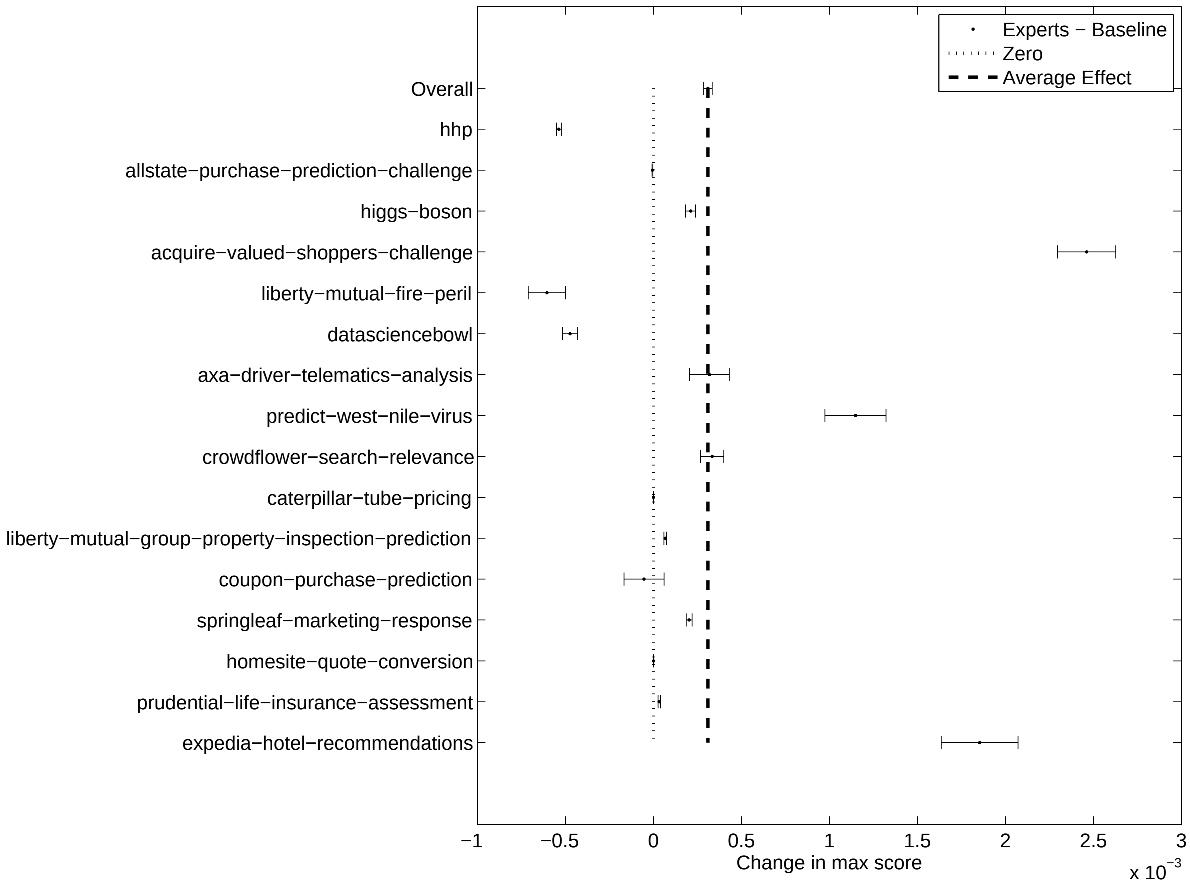


Figure A.7: Change in the Maximum Score when Comparing the Case of a Restricted Participation Contest with 90 Percent of the Teams—All of Them of the Highest Player Type—Versus the Case with the Observed Number of Teams (Baseline)

Note: An observation is a contest. The brackets indicate a 95 percent confidence interval.

An Intuition for Prizes

At a given state s, a player is considering to make a new submission. The expected payoff of making the submission, when the player think this is the last time she is playing but there will arrive n submissions from rival players after she makes her submission is

$$\lambda \sum_{i=1}^{k} V_i P_i(n|s)$$

where λ is the probability that the submission arrives before the end of the contest, V_i is the prize that a player gets when finishes ranked in the i-th position and $P_i(n|s)$ is the probability of finishing in position i, conditional on the current state s, given that there will be n future rival submissions before the end of the contest.

A player makes a new submission when this benefit is above the cost. Therefore, the incentive to play depends on the distribution of prizes. Suppose that $\sum_{i=1}^{k} V_i = V$, so what matters for incentives is

$$\sum_{i=1}^{k} \alpha_i P_i(n|s),$$

where $\alpha_i \in [0,1]$, $\sum_{i=1}^{k} \alpha_i = 1$, which is equivalent to

$$P_1 + \sum_{i=2}^{k} \alpha_i [P_i(n|s) - P_1(n|s)].$$

The incentive to participate with k prizes compared to a single prize for the winner then is larger when

$$\sum_{i=2}^{k} \alpha_i [P_i(n|s) - P_1(n|s)] > 0.$$

The difference in probability in our model depends on several pieces: (1) the distribution of types; (2) the number of players expected to arrive in the future; (3) the current state. Therefore, given multiple prizes may encourage or discourage participation. Notice, however, when $|P_i(n|s) - P_1(n|s)|$ is small the effect on incentives is also small. For this reason, the effect of one versus multiple prizes is not (empirically) very large. This difference is smaller when n is large, so we expect a small effect of prizes at the beginning of the contest and a larger effect towards the end of the contest, conditional on the state.