Races or Tournaments? [PRELIMINARY AND INCOMPLETE]*

Andrea Blasco Kevin J. Boudreau Karim R. Lakhani
Michael Menietti

Last updated: 25 January, 2017

Abstract

A wide range of economic and social situations are decided by either a race or a tournament. In either situations, agents choose whether and how much to exert some costly effort to increase the probability of being awarded a prize under the uncertainty about the actions of other agents. While a tournament yield outcomes greater than those of a race, the latter prevents unnecessary costs due to an excess of participation. We examine this trade-off empirically. We report the results of a field experiment conducted online where we compare the outcomes — efforts, quality, and diversity of outputs — of three alternative competitive situations motivated by theory: the race, the tournament, and the tournament with a quality requirement.

JEL Classification: xxx; xxx; xxx.

Keywords: xxxx; xxxx xxxx.

*Blasco: Harvard Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (email: `ablasco@fas.harvard.edu`).

Contents

1 Introduction

Organizations use often prizes as incentive schemes for innovation. Inside firms, contests are useful tools for exploiting and expanding the internal innovative capacity of the company. Contests are also sponsored by various institutions as a means to source innovative ideas from outside the boundaries of the organization, including the members of large online communities of practitioners.[1] And a growing number of platforms like Challenge.gov, Innocentive, Topcoder, and XPRIZE, have become available to help organize and connect these institutions with potential participants.

Typically, the objective of the contest designer is to maximize the quality and probability of a successful innovation while minimizing the time it takes to complete the project. Striking a balance between these two desirable but incompatible features it is often difficult. The lack of knowledge about the individual costs and technical feasibility of the innovation may force contest designers to make tough choices under considerable uncertainty. A growing literature on contest design offers insights on how to make better choices. However, while researchers have examined various aspects of the management of this trade-off, the optimal design remains a wide open issue.

Our contribution here is to focus on one key factor of design that is the choice of the way participants compete. In general, there are two extreme forms of competition. One is the *race* where the first to finish the innovation project wins. The second is the *tournament* where the best to finish the project within a given unit of time wins. To fix ideas, imagine a government designing an innovation contest to find a solution to a threat to the public health, such as the UK government with respect to the problem of antibiotic resistance.[2] To minimize the risks that the threat will materialize before a solution is actually found, one option is to design a contest with a tight deadline after which the government will rank order the received solutions and pick the most effective one, awarding a prize to the winners. Alternatively, the government can design a contest with a minimum desired effectiveness threshold and award a prize to the first solution that achieves or goes beyond the threshold. For a given prize amount, both approaches have specific advantages and limitations. If the prize to time ratio is too low, incentives in the tournament competition may not deliver high quality solutions. At the same time, in a race, participants have no incentives to exceed the threshold. Understing the conditions to choose one competition or the other is the main objective of this paper.

In this study, we design and execute a field experiment to compare behavior in a race

[1] In the United States, for instance, the government routinely sponsors open calls for innovation tackling important issues of public policy (public health, education, environment protection).

[2] This example is taken. . .

and a tournament setting. We proceed in two ways. First, we generalize an incomplete information contest model (Moldovanu and Sela, 2001) to enable a direct comparison of equilibrium behaviors under both the race and the tournament within a single theoretical framework. Then, we collect data from a field experiment to test some of the implications of the theory, and provide policy recommendations.

The field experiment was conducted xxxx. We worked together with researchers from the United States National Health Institute (NIH) and the Scripps Research Institute (SCRIPPS) to select a challenging problem for the contest. The selected problem was based on an algorithm called BANNER built by NIH (Leaman et al., 2008) that uses expert labeling to annotate abstracts from a prominent life sciences and biomedical search engine, PubMed, so disease characteristics can be more easily identified. The goal of the programming competition was to improve upon the current NIH's system by using a combination of expert and non-expert labeling, as described by Good et al. (2014). The competition was hosted online on the platform Topcoder.com (about 1M registered users in 2016). Submissions were made online. The top submissions were awarded a monetary prize ranging between $5000 to $100 for a total prize pool of $40,000.

Competitors were randomly assigned to virtual rooms of 10 or 15 people. These virtual rooms were then randomly assigned to one of three different competitive settings: a race, a tournament, and a tournament with a "reserve score," which is the lowest acceptable by the platform for a submission to be awarded a prize.

We find that xxxxx participation in the tournament is xxx compared to the race the reserve.

We also find that xxxx submission are quicker in a race, whereas are equally distributed at the end of the competition in the the tournament and in the tournament with quality requirement.

Another interesting finding is that xxxxx No evidence trade-off between a race and a tournament in terms of higher scores vs faster submissions. We do find that scores are higher in the tournament but we do not find a strong trade-off in the sense that race had comparable good quality solutions than the tournament.

An extensive literature has indicated the main theoretical reasons why contests are to be preferred to other forms of incentives. Typically, contests reduce monitoring costs [xxx], incentivize production with common risks [xxx], and deal with indivisible rewards [xxxx], among others. Much work has been done investigating several aspects of contest design, including the optimal prize structure [XXX, xxxx, xxxx], number of competitors [XXX, XXX], and imposing restrictions to competition such as minimum effort requirements [XXX, XXX].

Also, a great deal of theoretical models of races and tournaments have been developed and applied to a wide range of economic situations including patent races, arms races, sports, the mechanism of promotions inside firms, sales tournaments, xxx, and xxxx. A typical finding is that contests may lead to over-expenditure of effort with participants exerting more effort than the value of the prize. This problem happens when the contest designer does not capture the entire aggregate efforts which exceeds the value of the prize. In an innovation contest, xxxx we do not care. Another problem is that of a misallocation of the prize. This happens when participants have different evaluation for winning the prize and the contest is not xxxx (perfect discrimination). These issues have been discussed in the past. Here we take the perspective of the principal. Leaving aside misallocation, over-expenditure can be an issue also for the contest designer. For example, a manager care about the effort wasted xxxx working on a problem. Part is also with respect to the time spent, assuming workers can process one task at a time.

3.1.1 Excess of effort 3.1.2 Excess of participation

The dynamics. Harris and Vickers (1987), Grossman and Shapiro (1987) are classical models of dynamic patent races where the interest is how firms compete for a patent. Bimpikis et al. (2014) looks at the problem of how to design an information structure that is optimal when the contest is a race and innovation is uncertain (encouragement and competition effect). In the laboratory, Zizzo (2002) finds poor support to predictions of dynamic xxxx. In general we do not know much about the dynamic aspect of contests.

3.2 Duality

The duality. As pointed out by Baye and Hoppe (2003), many of these models of tournament and race competitions are specific cases of a more general "contest games." And sometimes it is possible to design one or the other in a way to exploit a "duality." In other words, in theory, a competition can be designed as a tournament to do xxx or as a race to do xxx. While theoretically very useful, how to exploit this duality in practice remains largely unknown. Lack of data. As before, xxxx. The main challenge is self-selection. The answer to this optimal design question relates to the cost function of agents with respect to "time" and to "effort." It is hard to say which solution is better. However, it is easier to tell whether you should have one prize or multiple prizes.

In a tournament, this type of preferences can be satisfied by fixing a deadline. Say time within which competitors are asked to provide their efforts. However, assuming competitors have costs from making less time in performing a task and there complementarities in costs, increasing the deadline in a tournament is similar to raising the marginal cost for everyone, which might not be an optimal solution. In a race, by contrast, increasing the deadline will affect entry but, conditional on entry, the time to complete the task will always be less than the deadline. Which means that those with low costs will be mostly affected by the deadline, whereas xxxx. Which may be a superior choice than the tournament.

2 Literature

This paper is related to the contest theory literature Dixit (1987) Baye and Hoppe (2003), Parreiras and Rubinchik (2010), Moldovanu and Sela (2001), Moldovanu and Sela (2006), Siegel (2009), Siegel (2014). It also relates to the literature on innovation contests Taylor (1995), Che and Gale (2003). And the personnel economics approach to contests Lazear and Rosen (1981), Green and Stokey (1983), Mary et al. (1984).

Empirically, Dechenaux et al. (2014) provide a comprehensive summary of the experimental literature on contests and tourments. Large body of empirical works have focused on sports contests Szymanski (2003). More recently, inside firms (xxx) and online contest (xxxx).

This paper is also related to the econometrics of auctions Paarsch (1992), Laffont et al. (1995), Donald and Paarsch (1996) and more recently Athey et al. (2011), Athey and Haile (2002), and Athey and Haile (2007).

3 The model

We generalize the *contest game* introduced by Moldovanu and Sela (2001) to a situation where players allocate their effort along two dimensions: performance and time. A (generalized) contest game is an N player game with asymmetric information. Players move simultaneously to maximize the expected utility. Each player $i = 1, 2, ..., N$ selects a performance variable y_i and a timing t_i, both being nonnegative numbers. These variables can be thought of as the quality of the solution given to a problem and the time to write the code implementing such solution. Players incur a cost from effort given by the function

$$C_i(y_i, t_i) = \frac{1}{a_i} c(y_i, t_i) \tag{1}$$

with $c_y(0) \geq 0$, $c_y{}' > 0$, $c_t(d) \geq 0$, and $c_t{}' < 0$. The parameter a_i denotes the player i's ability, which is privately observed at the beginning of the game. Abilities are drawn from a common distribution F that is continuous on the semi-infinite interval $[0, \infty)$ (e.g., the exponential distribution).

Let r_i denote the rank position of a player i relative to the $N-1$ others. The top K players (e.g., $r_i \leq K$) are awarded a prize of value $V_1 > V_2 > ... > V_K$. A player's probability of winning a prize is given by the function $p_i(y_1, ...y_N, t_1, ..., t_N)$.

The goal for each player is If the timing is above a given deadline $d > 0$ or the performance is below a certain level $q > 0$, the player gets zero utility. Otherwise, the player is given a rank based on his performance and timing. A contest is xxxx. Ranked .

Let d denote a given time unit and let q denote a given quality level.

Definition. A tournament is a contest game where players are ranked by their performance level y_i when they complete the project within the time unit (i.e., $t_i \leq d$).

Definition. A race is a contest game where players are ranked by their timing provided that the performance is above a minimum level $y_i \geq q$.

In both cases, if they xxxx they get no prizes (i.e., a rank $r > K$).

In a tournament, the agent having achieved the highest output quality within the deadline gets the first prize, the agent having achieved the second highest output quality gets the second prize, and so on. In a race, by contrast, the first agent to achieve an output quality of at least $\bar{y}$ within the deadline wins the first prize, the second to achieve the same target gets the second prize, and so on.

Each agent is risk neutral and faces the following decision problem

$$\text{maximize} \quad \textstyle\sum_{j=1}^{k} \Pr(\text{ranked } j\text{'th}) V_j - \frac{1}{a_i} C(y_i, t_i).$$

3.1 Equilibrium

We provide here the symmetric equilibrium with one prize and $n > 2$ agents. In appendix XXX, we provide a general formula for $k > 2$ prizes.

Let $y_{1:n} < y_{2:n} < ... < y_{n:n}$ denote the order statistics of the y_j's for every $j \neq i$ and let $F_{Y_{r:n}}(\cdot)$ and $f_{Y_{r:n}}(\cdot)$ denote the corresponding distribution and density for the r'th order statistic.

Proposition 1. *In a tournament, the unique symmetric equilibrium of the model gives, for every $i = 1, ..., n$, the optimal time to completion $t^*(a_i)$ equal to the deadline d and the optimal output quality $y^*(a_i)$ as*

$$y^*(a_i) = V_1 \int_{a_i}^{\infty} f_{Y_{n:n}}(z) dz$$

if $a_i \geq \underline{a}$ (see Moldovanu and Sela, 2001), and equal to zero otherwise.

An important property of xxx is that $y^*(a_i)$ has its upper bound in xxx and lower bound in xxxx. Also, equilibrium output quality is monotonic increasing in the agent's ability (see Moldovanu and Sela, 2001). Thus, for every $i = 1, ..., n+1$, the equilibrium expected reward depends only on the rank of his ability relative to the others. Using $F_{A_{r:n}}$ to denote the distribution of the r'th order statistic of abilities gives

$$F_{A_{n:n}}(a_i)V_1 - C(y_i^*, d, a_i).$$

Hence, by setting to zero and solving for the ability, gives the marginal ability $\underline{a}$ as

$$\underline{a} = h(n, V, F_A, C, d).$$

Corollary 1. *Equilibrium behavior in a race*

3.1.1 Contest designer's problem

The sponsor of the contest chooses the rules of the competition including prize structure $\{V_j\}_{j=1}^k$, deadline d, target quality q, and competition format (race or tournament). The sponsor maximizes an objective function that is the sum of total quality $Y = \sum_{i=1}^{n+1} Y_i$, time spent $T = \sum_{i=1}^{n+1} T_i$ and prizes paid $V = \sum_{j=1}^{k} p_j V_j$ (with $p_j = 1$ if the prize is awarded and $p_j = 0$ otherwise). Hence, the problem faced by the sponsor is

$$\text{maximize} \quad \int Y - \tau \mathbf{E}T - \mathbf{E}V$$

with the intensity of preferences towards time weighted by $c_t \geq 0$.

3.2 Structural econometric model

4 The experimental design

The field experiment was conducted between March 2 and 16, 2016. The context of the experiment was an online programming contest. In an online programming contest, participants compete to write source code that solves a designated problem. These contests are quite common and xxxx either as a tournament or a race competition.

The contest was hosted on the online platform Topcoder.com. Since its launch in 2001, Topcoder.com administers on a weekly basis several competitive programming contests for thousands of competitors from all over the world. Typical assigned problems are data science problems (e.g., classification, prediction, natural language processing) that demand some background in machine learning and statistics. All Topcoder members (about 1M registered users in 2016) can compete and attain a "rating" that provides a metric of their ability as contestants. Other than attaining a rating, the competitors having made the top five submissions in a competition are typically awarded a monetary prize the extent of which depends on the nature and complexity of the problem but is generally between $5,000 and $20,000.

In this study, we worked together with researchers from the United States National Health Institute (NIH) and the Scripps Research Institute (SCRIPPS) to select a challenging problem for the experimental programming competition. The selected problem was based on an algorithm called BANNER built by NIH (Leaman et al., 2008) that uses expert labeling to annotate abstracts from a prominent life sciences and biomedical search engine, PubMed, so disease characteristics can be more easily identified. The goal of the programming competition was to improve upon the current NIH's system by using a combination of expert and non-expert labeling, as described by Good et al. (2014).

The competition was announced on the platform and to all community members via email. A preliminary online registration was required to enroll in the competition, which resulted in 340 pre-registered participants. Among the pre-registered members, we selected the 299 who had registered to a programming contest at least once before the present contest. This choice was to ensure that participants were xxxx.

Participants were then randomly assigned to separate groups of 10 or 15 people. In each of these groups, contestants were given access to a "virtual room" that is a private web page listing handles of the other participants of the group, a leaderboard updated regularly during the competition, and a common chat that they can use to ask clarifying questions (visible to everyone in the group) with respect to the problem at hand.

A problem statement containing a full description of the algorithmic challenge, the rules of the game, and payoffs was published at the beginning of the submission phase. The submission phase was of 8 days in which participants could submit their computer programs. Each submission was automatically scored and feedback in the form of preliminary scores was published regularly on the website via the leaderboard.

Groups were randomly assigned to one of three different competitive settings: a race, a tournament, and a tournament with a *reserve target*, which is the lowest acceptable score by the platform for a submission to be awarded a prize.

The experimental design is summarized by the Table XXXX.

In all groups, the first placed competitor was awarded a prize of $1,000, and an additional, consolatory prize of $100 was awarded to the second one.

In a race competition, however, the first to achieve a score equal to xxxx was placed

Table 1: Experimental design

	Large	Small
Race	60	39
Tournament	60	40
Reserve	60	40

first. The level was chosen xxxx.

In a tournament, xxxx.

Finally, in a tournament with reserve, xxxx.

Additional grand prizes of xxxx were awarded to the top xxx in every treatment.

4.1 Data

The bulk of our data comes from the online Topcoder's profile of each participant. This profile typically includes information of when the member registered to the platform, the current rating in a variety of different competitions, the number of past competitions, and so on. Additional demographic information, was collected via a pre-registration survey where competitors were asked to state their gender, age, geographic origin, etc. Participants were als also asked a self-reported measure of risk aversion [xxx] and to forecast how many hours they expected to compete in the next few days of the challenge(the exact question was: "looking ahead xxxx").

Finally, we also asked participants to respond to a survey at the end of the submission phase. In this final survey, they were asked to look back and tell us their best estimate of the time spent working on the problem. Also, we gathered comments on the xxx. And questions such as xxxx.

Table XXX summarizes the data.

[This goes into the data section:] A total of 299 competitors signed-up to take part in the challenge. They were all xxxx members of the platform with between 52.542 and 770.548 weeks as registered members. In terms of skill ratings, the distribution was highly skewed with competitors in the highest 90th percentile having participated in 24 more competitions than those in the 10th percentile. Likewise skills as measured by the individual ratings, if there was one, had a skewed distribution with 1034 higher points than those in the 10th percentile; see Figure **??**.

Table 2: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Algo rating	299	1,051.0	730.0	0	2,958
Algo competitions	299	40.6	57.7	0	338
Algo registrations	299	45.7	64.3	0	365
MM rating	205	1,322.0	425.0	593	3,071
MM competitions	299	7.2	11.8	0	91
MM registrations	299	17.6	23.0	1	161
Time zone	279	2.1	5.1	−8.0	10.0
Latitude	279	36.7	19.0	−42.8	59.9
Longitude	279	25.2	77.7	−122.0	149.0
Risk aversion	284	6.4	2.2	1	10

5 Results

In the eight-day submission period, we collected a total of 1759 submissions made by 86 participants. The frequency distribution of submissions was rather skewed with participants in the 90th percentile making 50 more submissions than those in the 10th percentile.

```
##
##          Race Tournament Reserve
##   FALSE   73          67      73
##   TRUE    26          33      27
```

Participation across treatments was higher in the Tournament treatment (33 percent), followed by the Tournament w/reseve (27 percent), and the Race treatment (26.263 percent). However, these differences are not statistically significant (using a Fisher's Exact Test for Count Data gives a p-value of 0.526)

```
##
##          Race Tournament Reserve
##   FALSE   84          81      85
##   TRUE    15          19      15
```

Similar results are found when we consider only those who made submissions above a score of xxxx. Participation across treatments was higher in the Tournament treatment (19 percent), followed by the Tournament w/reseve (15 percent), and the Race treatment (15.152 percent). However, these differences are not statistically significant (using a Fisher's Exact Test for Count Data gives a p-value of 0.699)

Regarding to the frequency of submissions per participant, the median count is larger in the race and in the tournament w/reserve (12 and 15 respectively) compared to the Tournament (a median of 6 submissions). However, a Kruskal-Wallis rank sum test fails to reject the null hypothesis (p=0.794) that at least one treatment was different in the location of the frequency distribution of the submissions.

Concerning the distribution of the scores on the last submission, the median final scores was higher in the Tournament w/reserve treatment (0.808), followed by the Tournament (0.805), and the Race treatment (0.799). As before, however, these differences are not statistically significant (a Kruskal-Wallis rank sum test gives a p-value of 0.577).

Although we do not find significant differences through an univariate analysis, it is possible that differences will be xxx in a multivariate analysis. Adding controls can indeed reduce noise and improve precisions of our estimates.

Assuming the decision was independent, to explore the determinants of participation:

$$Pr(y = 1) = G(\text{Rating}_i + \text{Experience}_i + \text{T}_i) \tag{2}$$

where $G()$ is logistic.

```
##
## ==========================================================
##                                   Dependent variable:
##                              ------------------------------
##                                         submit
##                                 (1)        (2)        (3)
## ----------------------------------------------------------
## poly(mmevents, deg = 3)1 12.200***
##                           (4.120)
##
## poly(mmevents, deg = 3)2   1.210
##                           (6.130)
##
## poly(mmevents, deg = 3)3  8.110*
##                           (4.740)
##
## poly(mmevents, deg = 2)1             4.690*     4.980*
##                                     (2.540)    (2.590)
##
```

```
## poly(mmevents, deg = 2)2           -0.212    -0.077
##                                    (2.470)   (2.530)
##
## poly(mmrating, deg = 2)1          10.100*** 8.530***
##                                    (3.120)   (3.260)
##
## poly(mmrating, deg = 2)2           -0.942    -1.500
##                                    (2.380)   (2.430)
##
## lat                                          0.021**
##                                              (0.009)
##
## long                                          0.003
##                                              (0.002)
##
## Constant                 -0.961*** -1.020*** -1.900***
##                           (0.138)   (0.143)   (0.405)
##
## -------------------------------------------------------
## Observations                299       299       279
## Log Likelihood           -166.000  -164.000  -150.000
## Akaike Inf. Crit.         341.000   337.000   315.000
## =======================================================
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: submit
##
## Terms added sequentially (first to last)
##
##
##                          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                       278        334
## poly(mmevents, deg = 2)  2    17.87        276        316  0.00013
```

```
## poly(mmrating, deg = 2)  2     8.89        274         308  0.01172
## lat                      1     4.96        273         303  0.02589
## long                     1     1.85        272         301  0.17347
```

This result does not seem to correlate well with the competitor's experience or skills, as the Pearsons's correlation coefficient between the count of past competitions or the rating and the count of submissions is positive but generally low; see Table XXX. Thus, differences in submissions appear idiosyncratic and perhaps related to the way to organize the work rather than systematically associated with underlying differences in experience or skills.

The timing of submissions was rather uniform during the submission period with a peak of submissions made in the last of the competition. (explain more)

```
#scores$submax <- ave(subs$sub_id, subs$handle, FUN=max)
#par(mfrow=c(2, 1), mar=c(4,4,2,2))
#plot(subid==1 ~ as.POSIXct(subts), data=scores, type='h', yaxt='n'
#    , xlab='', ylab='', main='Dispersion time first submission')
#plot(subid==submax ~ as.POSIXct(subts), data=scores, type='h'
#    , yaxt='n', xlab='', ylab='', main='Dispersion time last submission')
```

Scores: xxxx

5.1 Treatment differences

Difference in participation by treatments are show in Table XX.

```
Fisher's Exact Test for Count Data
```

data: tab p-value = 0.5 alternative hypothesis: two.sided

We find no differences in the room size.

```
Fisher's Exact Test for Count Data
```

data: tab p-value = 1 alternative hypothesis: true odds ratio is not equal to 1 95 percent confidence interval: 0.569 1.691 sample estimates: odds ratio 0.985

Ex-post

Timing: early vs late

Using a Chi-square test of independence, we find no significant differences in participation rates associated with the assigned treatments (p-value: 1); see Table XX.

Further, we model participation rates as a logistic regression. We use a polynomial of third degree for the count of past competitions to account for non-linear effects of experience; and we use an indicator for whether the competitor had a win or not. Also, taking into account differences in ability, participation rates are not significantly different.

5.2 Estimation results

Participation to the competition by treatment is shown in Figure **??**. Participation here is measured by the proportion of registered participants per treatment who made any submission during the eight-day submission period. Recall that competitors may decide to enter into the competition and work on the problem without necessarily submitting. In a tournament, for example, competitors are awarded a prize based on their last submission and may decide to drop out without submitting anything. However, this scenario seems unlikely. In fact, competitors often end up making multiple submissions because by doing so they obtain intermediate feedback via preliminary scoring (see Section XXX for details). In a race, competitors have even stronger incentives to make early submissions as any submission that hits the target first wins.

`Table xxx`

We find that the propensity to make a submission is higher in the Tournament than in the Race and in the Tournament with reserve, but the difference is not statistically significant (a Fisher's exact test gives a p-value of xxxxx). As discussed in Section XXX, we may not have enough power to detect differences below 5 percentage points. However, we find the same not-significant result in a parametric regression analysis of treatment differences with controls for the demographics and past experience on the platform; see Table **??**. Adding individual covariates reduces variability of outcomes, potentially increasing the power of our test. In particular, Table **??** reports the results from a logistic regression on the probability of making a submissions. Column 1 reports the results from a baseline model with only treatment dummies. Column 2 adds demographics controls, such as the age, education, and gender. Column 3 adds controls for the past experience on the platform. Across all these specifications, the impact of the treatment dummies (including room size) on entry is not statistically significant.

5.3 Simulation results

6 Empirical analysis

6.1 Estimation results

Participation to the competition by treatment is shown in Figure

$$fig : entry$$

. Participation here is measured by the proportion of registered participants per treatment who made any submission during the eight-day submission period. Recall that competitors may decide to enter into the competition and work on the problem without necessarily submitting. In a tournament, for example, competitors are awarded a prize based on their last submission and may decide to drop out without submitting anything. However, this scenario seems unlikely. In fact, competitors often end up making multiple submissions because by doing so they obtain intermediate feedback via preliminary scoring (see Section XXX for details). In a race, competitors have even stronger incentives to make early submissions as any submission that hits the target first wins.

`Table xxx`

We find that the propensity to make a submission is higher in the Tournament than in the Race and in the Tournament with reserve, but the difference is not statistically significant (a Fisher's exact test gives a p-value of xxxx). As discussed in Section XXX, we may not have enough power to detect differences below 5 percentage points. However, we find the same not-significant result in a parametric regression analysis of treatment differences with controls for the demographics and past experience on the platform; see Table

$$entry$$

. Adding individual covariates reduces variability of outcomes, potentially increasing the power of our test. In particular, Table

$$entry$$

reports the results from a logistic regression on the probability of making a submissions. Column 1 reports the results from a baseline model with only treatment dummies. Column 2 adds demographics controls, such as the age, education, and gender. Column

3 adds controls for the past experience on the platform. Across all these specifications, the impact of the treatment dummies (including room size) on entry is not statistically significant.

6.2 Simulation results

References

Athey, Susan and Philip A Haile, "Identification of standard auction models," *Econometrica*, 2002, *70* (6), 2107–2140.

_ and _ , "Nonparametric approaches to auctions," *Handbook of econometrics*, 2007, *6*, 3847–3965.

_ , Jonathan Levin, and Enrique Seira, "Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions*," *The Quarterly Journal of Economics*, 2011, *126* (1), 207–257.

Baye, Michael R and Heidrun C Hoppe, "The strategic equivalence of rent-seeking, innovation, and patent-race games," *Games and Economic Behavior*, 2003, *44* (2), 217–226.

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

Che, Yeon-Koo and Ian Gale, "Optimal design of research contests," *The American Economic Review*, 2003, *93* (3), 646–671.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta, "A survey of experimental research on contests, all-pay auctions and tournaments," *Experimental Economics*, 2014, pp. 1–61.

Dixit, Avinash Kamalakar, "Strategic Behavior in Contests," *The American Economic Review*, 1987, *77* (5), 891–98.

Donald, Stephen G and Harry J Paarsch, "Identification, estimation, and testing in parametric empirical models of auctions within the independent private values paradigm," *Econometric Theory*, 1996, *12* (03), 517–567.

Good, Benjamin M, Max Nanis, CHUNLEI Wu, and ANDREW I Su, "Microtask crowdsourcing for disease mention annotation in PubMed abstracts.," in "Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing" NIH Public Access 2014, pp. 282–293.

Green, Jerry R and Nancy L Stokey, "A Comparison of Tournaments and Contracts," *The Journal of Political Economy*, 1983, *91* (3), 349–364.

Grossman, Gene M and Carl Shapiro, "Dynamic R&D Competition," *Economic Journal*, 1987, *97* (386), 372–87.

Harris, Christopher and John Vickers, "Racing with uncertainty," *The Review of Economic Studies*, 1987, *54* (1), 1–21.

Laffont, Jean-Jacques, Herve Ossard, and Quang Vuong, "Econometrics of First-Price Auctions," *Econometrica*, 1995, *63* (4), 953–80.

Lazear, Edward P and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 1981, *89* (5), 841–864.

Leaman, Robert, Graciela Gonzalez et al., "BANNER: an executable survey of advances in biomedical named entity recognition.," in "Pacific symposium on biocomputing," Vol. 13 2008, pp. 652–663.

Mary, O'Keeffe, W Kip Viscusi, and Richard J Zeckhauser, "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, 1984, *2* (1), 27–56.

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

_ **and** _ , "Contest architecture," *Journal of Economic Theory*, 2006, *126* (1), 70–96.

Paarsch, Harry J, "Deciding between the common and private value paradigms in empirical models of auctions," *Journal of econometrics*, 1992, *51* (1-2), 191–215.

Parreiras, Sérgio O and Anna Rubinchik, "Contests with three or more heterogeneous agents," *Games and Economic Behavior*, 2010, *68* (2), 703–715.

Siegel, Ron, "All-Pay Contests," *Econometrica*, 2009, *77* (1), 71–92.

_ , "Contests with productive effort," *International Journal of Game Theory*, 2014, *43* (3), 515–523.

Szymanski, Stefan, "The Economic Design of Sporting Contests," *Journal of Economic Literature*, 2003, *41* (4), 1137–1187.

Taylor, Curtis R, "Digging for Golden Carrots: An Analysis of Research Tournaments," *The American Economic Review*, 1995, *85* (4), 872–90.

Zizzo, Daniel John, "Racing with uncertainty: a patent race experiment," *International Journal of Industrial Organization*, 2002, *20* (6), 877–902.