Races or Tournaments? [PRELIMINARY AND INCOMPLETE]*

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

Last updated: 31 January, 2017

Abstract

Contests are often used as incentive schemes to foster innovation. The typical goal of contest designers is to maximize quality while minimizing the time it takes to achieve the innovation. This situation leads to a difficult choice of design made under considerable uncertainty. In this study, we investigate one key aspect of this decision that is the way participants compete. Two extreme forms of competition are considered: the race, where the first to achieve the innovation wins, and the tournament, where the timing is not important. We develop a model to characterize under what conditions contest designers should go for one or the other approach. Then, we report the results of a field experiment conducted to compare the outcomes of three alternative competitive situations motivated by theory: the race, the tournament, and the tournament with a quality requirement. We find that outcomes in a race are of comparable quality, but are supplied faster. Based on our model, we also show what would be optimal to do under several simulated counterfactual situations.

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 often use contests as incentive schemes to foster innovation. In an innovation contest, participants compete for prizes by investing time and energy in an innovation project. The typical aim of the contest designer is to maximize quality while minimizing the time to complete the innovation. Balancing between these two desirable but incompatible features it is often difficult. The lack of knowledge about individual costs and technical feasibility of the innovation may force contest designers to make their choices under considerable uncertainty. A growing literature on contest design offers insights on how to make better resolutions. However, while researchers have examined various aspects of managing this trade-off, the optimal design is still largely unknown.

Here, we investigate one key aspect of contest design concerning the way participants compete. There are two extreme forms of competition that we consider. One is the *race* where the first to finish an innovation project wins. The second is the *tournament* where the best finished project 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 on the problem of antibiotic resistance.[1] To minimize the risks that the threat will materialize before a solution is found, one choice is a tournament competition with a tight deadline for participants to provide their solutions. Alternatively, the government can use a race competition awarding a prize to the first participant that achieves, or goes beyond, a minimum effectiveness threshold. Fixed the prize, both approaches have specific advantages and limitations. If the duration of the competition is too low, a tournament may give insufficient incentives resulting in inadequate solutions. In a race competition, instead, timing is not an issue but participants have no incentives to exceed the minimum threshold.

In the present study, we shed light on the conditions under which contest designers should choose between a race and a tournament competition. We proceed in two ways. First, we develop a contest model that encompasses both the race and the tournament in a single framework. Exploring the duality of the model, we compare equilibrium behaviors under both regimes and characterize the optimal choice (i.e., the setup that maximizes the utility of the contest designer). Then, we design and execute an experiment to test the implications of the theory in the field, providing policy recommendations.

Regarding to the modeling, we adapt the contest model introduced by Moldovanu and Sela (2001). Contests have an all-pay structure by which participants pay an immediate cost for an uncertain future reward. We generalize by allowing participants to choose

[1] This example is taken. . .

the timing and quality of the innovation at once. This decision is made under the uncertainty of the costs of the rivals, which are privately observed by the agents. The contest designer is modeled as an extra agent with preferences for both time and quality. Following the analysis of the model model, we show that the optimal design depends on the number of participants and the concavity of their cost function.

The field experiment was conducted at the end of 2016. The context of the experiment was an online programming competition. In a programming competition, participants compete writing source code that solves a given problem for winning a monetary prize. 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). Top submissions were awarded monetary prizes ranging between $100 to $5000 for a total prize pool of more than $40,000.

Our intervention consisted in sorting at random participants into independent 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 score 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.]

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).

An extensive literature has discussed the reasons why contests are sometimes preferred to other forms of incentives (e.g., individual contracts). Typically, contests reduce monitoring costs [xxx], incentivize production with common risks [xxx], and deal with indivisible rewards [xxxx], among others. While there is not much debate on why contests should be used, the issue of how to effectively design and deploy a contest still attracts much research.

Several aspects of contest design have been investigated, 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 [xxx], arms races [xxx], sports [xxx], the mechanism of promotions inside firms [xxxx], sales tournaments [xxxx], etc.

Harris and Vickers (1987), Grossman and Shapiro (1987) investigate the dynamics issues 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.

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.

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 contest game is an N player game with asymmetric information. Players move simultaneously to maximize the expected utility. Each player $i = 1, 2, ..., N$ chooses a performance variable y_i and a timing t_i, both being nonnegative numbers. These can be thought of as the quality of the solution given to a given data problem and the time to write the code implementing such solution.

Players incur a cost from effort given by the function

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

with $c_y(0) \geq 0$, $c_y{}' > 0$, $c_t(d) \geq 0$, and $c_t{}' < 0$. The function $\gamma(\cdot)$ is a cost-shift that depends on the player i's ability a_i, which is privately observed at the beginning of the game. Although private, it is common knowledge that abilities are drawn from a common distribution F that is continuous on the semi-infinite interval $[0, \infty)$.

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

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

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.

Table 1: Experimental design

	Large	Small
Race	60	39
Tournament	60	40
Reserve	60	40

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 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.

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 **??**.

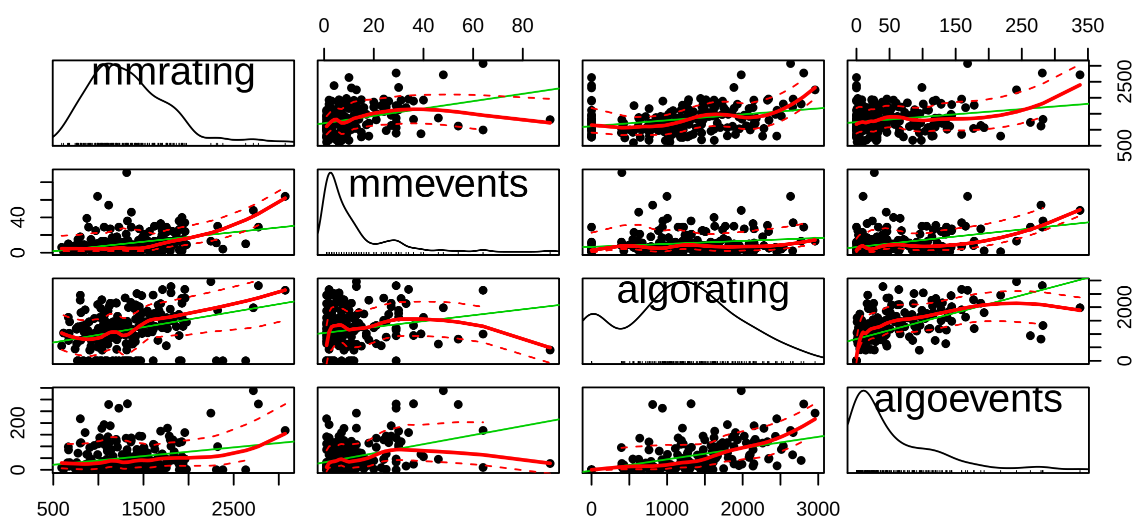


Table 2: Descriptive statistics

Variable	Response Category	Frequency	P-value
country_name	India	36	1.000
	China	31	
	United States	31	
	Japan	27	
	Russia	26	
	(Other)	148	
age	<20 years old	18	0.893
	20-25 years old	99	
	26-30 years	82	
	31-40 years	67	
	41-50 years	21	
	51 years and above	12	
gender	Female	20	0.672
	Male	279	
educ	PhD	31	0.454
	High School	37	
	Master of Arts	111	
	Bachelor	120	
plang	C#	33	0.640
	C++	134	
	Java	87	
	Other	14	
	Python	28	
	VB	3	

Notes: This table shows the frequency for each response category of five categorical variables for country of residence, age, gender, highest degree achieved, and preferred programming language. For each variable, a Pearson's Chi-squared test shows no evidence of dependence between the variable and the treatments. P-values are reported in the last column.

Table 3: 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	299	2.3	5.1	−8.0	10.0
Risk aversion	279	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).

```
##       2       4       6      11      13      19      29      30      31
##   0.684 138.976 141.476 139.105  46.955  30.184 189.441 181.222  35.447
```

```
##      33      41      42      43      44      49      51      59      60 
##  36.994  39.664  43.488  77.105  56.271 124.843  95.768  14.890  15.478 
##      72      77      82      87      90      93      94      96      97 
##   5.964 162.668  17.883 157.248  96.531 142.929 107.964  99.938 184.409 
##      98     100     104     106     109     110     119     121     122 
##  51.772  57.979  40.233  22.424  40.185  23.480  13.537   6.214  46.872 
##     131     137     142     145     146     147     150     151     152 
##  46.062  50.034  21.327  26.544 161.065   0.877  27.265  60.516  61.164 
##     158     160     163     169     171     172     174     178     183 
##  18.180  51.831   3.778  45.526   3.655 164.592  63.410  79.493 189.279 
##     193     196     198     202     207     209     212     213     215 
##  45.980 178.221  12.716  75.923  33.750 140.688   1.202  95.588   4.277 
##     220     224     225     228     230     236     239     247     251 
## 115.799 183.722  75.059   0.000  42.924  24.107  28.968 129.224  61.141 
##     254     266     274     275     276     277     278     280     285 
## 166.852 162.826  27.114  65.826   3.420  14.974 171.796 177.358   6.125 
##     289     291     292     293     299 
##  29.739  33.144 101.579  59.901 125.433
```

```
##       2       4       6      11      13      19      29      30      31 
## -190.40  -52.11  -49.61  -51.98 -144.13 -160.91   -1.65   -9.87 -155.64 
##      33      41      42      43      44      49      51      59      60 
## -154.09 -151.42 -147.60 -113.98 -134.82  -66.25  -95.32 -176.20 -175.61 
##      72      77      82      87      90      93      94      96      97 
## -185.12  -28.42 -173.21  -33.84  -94.56  -48.16  -83.13  -91.15   -6.68 
##      98     100     104     106     109     110     119     121     122 
## -139.32 -133.11 -150.86 -168.66 -150.90 -167.61 -177.55 -184.88 -144.22 
##     131     137     142     145     146     147     150     151     152 
## -145.03 -141.05 -169.76 -164.54  -30.02 -190.21 -163.82 -130.57 -129.93 
##     158     160     163     169     171     172     174     178     183 
## -172.91 -139.26 -187.31 -145.56 -187.43  -26.50 -127.68 -111.60   -1.81 
##     193     196     198     202     207     209     212     213     215 
## -145.11  -12.87 -178.37 -115.17 -157.34  -50.40 -189.89  -95.50 -186.81 
##     220     224     225     228     230     236     239     247     251 
##  -75.29   -7.37 -116.03 -191.09 -148.16 -166.98 -162.12  -61.86 -129.95 
##     254     266     274     275     276     277     278     280     285 
##  -24.24  -28.26 -163.97 -125.26 -187.67 -176.11  -19.29  -13.73 -184.96
```

```
##      289     291     292     293     299
## -161.35 -157.95  -89.51 -131.19  -65.66
```

Finally, let us focus on the timing of the first and last submission. xxxx

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.

Let's consider first a simple logistic model

```
##
## Call:
## glm(formula = submit ~ treatment, family = quasibinomial, data = races)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -0.895  -0.793  -0.781   1.489   1.635
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.0324     0.2295   -4.50  9.9e-06
## treatmentTournament   0.3242     0.3136    1.03     0.30
## treatmentReserve      0.0377     0.3224    0.12     0.91
##
## (Dispersion parameter for quasibinomial family taken to be 1.01)
##
##     Null deviance: 358.81  on 298  degrees of freedom
## Residual deviance: 357.49  on 296  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = submit ~ treatment + gender + educ + age + timezone +
##     plang, family = quasibinomial, data = races)
##
## Deviance Residuals:
```

```
##    Min      1Q  Median      3Q     Max
## -1.533  -0.840  -0.648   1.168   2.079
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -1.8533     0.9969   -1.86    0.064
## treatmentTournament     0.2977     0.3381    0.88    0.379
## treatmentReserve        0.0454     0.3491    0.13    0.897
## genderMale              0.3312     0.6178    0.54    0.592
## educHigh School         0.8063     0.6322    1.28    0.203
## educMaster of Arts      0.2427     0.4906    0.49    0.621
## educBachelor            0.5050     0.5141    0.98    0.327
## age20-25 years old     -0.3094     0.6757   -0.46    0.647
## age26-30 years          0.5366     0.7060    0.76    0.448
## age31-40 years          0.9886     0.7021    1.41    0.160
## age41-50 years          0.4718     0.8314    0.57    0.571
## age51 years and above   1.1899     0.9109    1.31    0.193
## timezone                0.0109     0.0279    0.39    0.695
## plangC++               -0.1932     0.4441   -0.44    0.664
## plangJava              -0.8340     0.4754   -1.75    0.080
## plangOther             -0.4482     0.7454   -0.60    0.548
## plangPython             0.2196     0.5838    0.38    0.707
## plangVB                 1.1734     1.3532    0.87    0.387
##
## (Dispersion parameter for quasibinomial family taken to be 1.06)
##
##     Null deviance: 358.81  on 298  degrees of freedom
## Residual deviance: 337.25  on 281  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = submit ~ treatment + algorating + mmevents + algoevents +
##     totalpayments, family = quasibinomial, data = races)
##
```

```
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.713  -0.771  -0.644   1.040   1.908
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -1.503410   0.327239   -4.59  6.5e-06
## treatmentTournament        0.372238   0.331949    1.12   0.2631
## treatmentReserve          -0.000123   0.343707    0.00   0.9997
## algorating                -0.000059   0.000228   -0.26   0.7957
## mmevents                   0.035712   0.013466    2.65   0.0084
## algoevents                 0.001907   0.003090    0.62   0.5377
## totalpayments1 - 599      -0.182154   0.477411   -0.38   0.7031
## totalpayments600 - 4500    0.226796   0.456516    0.50   0.6197
## totalpayments4500 - 37000  0.723311   0.426958    1.69   0.0913
## totalpayments>37000        0.466569   0.452361    1.03   0.3032
##
## (Dispersion parameter for quasibinomial family taken to be 1.03)
##
##     Null deviance: 358.81  on 298  degrees of freedom
## Residual deviance: 334.04  on 289  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = submit ~ treatment + algorating + mmevents + algoevents +
##     totalpayments + gender + educ + age + timezone + plang, family = qua
##     data = races)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.966  -0.759  -0.597   0.992   2.211
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                -2.13e+00  1.07e+00  -1.99    0.047
## treatmentTournament         3.50e-01  3.56e-01   0.98    0.327
## treatmentReserve           -5.13e-03  3.70e-01  -0.01    0.989
## algorating                 -3.52e-05  2.62e-04  -0.13    0.893
## mmevents                    3.50e-02  1.61e-02   2.17    0.031
## algoevents                  1.65e-03  3.34e-03   0.50    0.621
## totalpayments1 - 599       -2.12e-02  5.18e-01  -0.04    0.967
## totalpayments600 - 4500     2.56e-01  5.07e-01   0.50    0.614
## totalpayments4500 - 37000  7.25e-01   4.68e-01   1.55    0.123
## totalpayments>37000         3.60e-01  5.11e-01   0.70    0.482
## genderMale                  2.56e-01  6.51e-01   0.39    0.694
## educHigh School             8.29e-01  6.66e-01   1.24    0.215
## educMaster of Arts          1.39e-01  5.17e-01   0.27    0.788
## educBachelor                4.14e-01  5.43e-01   0.76    0.446
## age20-25 years old         -2.84e-01  7.17e-01  -0.40    0.692
## age26-30 years              3.45e-01  7.61e-01   0.45    0.651
## age31-40 years              6.57e-01  7.65e-01   0.86    0.391
## age41-50 years             -2.77e-01  9.49e-01  -0.29    0.770
## age51 years and above       5.57e-01  1.02e+00   0.55    0.584
## timezone                    1.17e-02  3.03e-02   0.39    0.700
## plangC++                   -4.17e-02  4.72e-01  -0.09    0.930
## plangJava                  -6.66e-01  4.99e-01  -1.34    0.183
## plangOther                 -2.35e-03  7.83e-01   0.00    0.998
## plangPython                 3.38e-01  6.13e-01   0.55    0.582
## plangVB                     7.60e-01  1.51e+00   0.50    0.615
##
## (Dispersion parameter for quasibinomial family taken to be 1.09)
##
##     Null deviance: 358.81  on 298  degrees of freedom
## Residual deviance: 320.83  on 274  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4

##
## ======================================================
##                              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*    4.990*
##                                     (2.540)   (2.590)
##
## poly(mmevents, deg = 2)2            -0.212    -0.087
##                                     (2.470)   (2.530)
##
## poly(mmrating, deg = 2)1           10.100*** 8.520***
##                                     (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)   (0.404)
##
## -----------------------------------------------------------
## 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
```

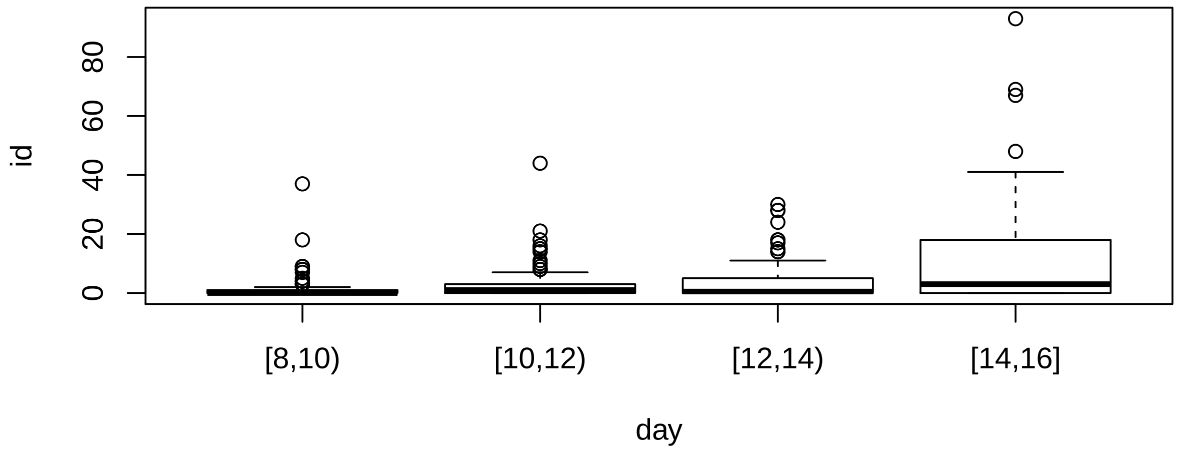
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: submit
##
## Terms added sequentially (first to last)
##
##
##                            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                         278        334
## poly(mmevents, deg = 2)  2    17.87        276        316  0.00013
## poly(mmrating, deg = 2)  2     8.89        274        308  0.01172
## lat                      1     4.97        273        303  0.02581
## long                     1     1.83        272        301  0.17578
```

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$submax <- ave(races.sub$id, races.sub$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')
```

Consider panel data!




```
##
## Call:
## glm(formula = id ~ day, data = subs.long)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -11.69   -4.00   -1.59    0.08   81.31
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.59       1.08    1.47     0.14
## day[10,12)      1.58       1.53    1.03     0.30
## day[12,14)      2.41       1.53    1.57     0.12
## day[14,16]     10.09       1.53    6.60  1.6e-10
##
## (Dispersion parameter for gaussian family taken to be 101)
##
##     Null deviance: 39401  on 343  degrees of freedom
## Residual deviance: 34190  on 340  degrees of freedom
## AIC: 2568
##
## Number of Fisher Scoring iterations: 2
```

```
##
## Call:
## glm(formula = id ~ day, family = quasipoisson, data = subs.long)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -4.834  -2.828  -1.785   0.023  14.939
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.466      0.342    1.36    0.174
## day[10,12)     0.689      0.419    1.65    0.101
## day[12,14)     0.921      0.404    2.28    0.023
## day[14,16]     1.993      0.365    5.47  8.9e-08
##
## (Dispersion parameter for quasipoisson family taken to be 16)
##
##     Null deviance: 4500.3  on 343  degrees of freedom
## Residual deviance: 3587.7  on 340  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
##
## Call:
## glm(formula = id > 0 ~ day, data = subs.long)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -0.733  -0.500   0.267   0.465   0.651
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3488     0.0521    6.70  8.7e-11
## day[10,12)    0.1860     0.0736    2.53    0.012
## day[12,14)    0.1512     0.0736    2.05    0.041
## day[14,16]    0.3837     0.0736    5.21  3.3e-07
```

```
##
## (Dispersion parameter for gaussian family taken to be 0.233)
##
##     Null deviance: 85.709  on 343  degrees of freedom
## Residual deviance: 79.279  on 340  degrees of freedom
## AIC: 481.4
##
## Number of Fisher Scoring iterations: 2
```

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.