Prize Allocation and Entry in Ideation Contests

Pavel Kireyev*

Harvard Business School

September 20, 2015

JOB MARKET PAPER

Abstract

Contests are a popular mechanism for the procurement of creative innovation. In marketing, firms often organize contests online, offering prizes to encourage competition and solicit high-quality ideas for ads, new products, and even marketing strategies from participants. I empirically investigate the impact of the number of prizes, prize amount and submission limit on participation and quality outcomes in ideation contests using data from a popular marketing crowdsourcing platform. I develop a structural model of participant entry and sponsor choice in contests with multiple prizes and heterogeneous participants. The structural model allows for multiple equilibria in the participant entry game and a flexible specification of participant information sets. Counterfactual simulations reveal the impact of design parameters on participation and quality outcomes: multiple prizes discourage stronger participants and encourage weaker participants but do not have a substantial impact on outcomes in contests that attract a large number of submissions; a larger prize increases expected total and maximum idea quality but may not substantially increase participation; a submission limit increases the number of entrants but reduces expected total and maximum idea quality. The results provide guidance for the optimal design of ideation contests.

Keywords: crowdsourcing; ideation; contest design; structural estimation; static discrete games; information structures; partial identification.

*Address: Morgan Hall, Soldiers Field Road, Boston, MA 02163, e-mail: *pkireyev@hbs.edu*. I thank my committee, Elie Ofek, Sunil Gupta and Ariel Pakes for their support. I thank Karim Lakhani, Robin Lee, Donald Ngwe, Daniel Pollmann, Thomas Wollmann and participants of the Harvard IO Lunch and the HBS Marketing Unit Seminar for helpful discussions. I thank James DeJulio and Eugene Retunsky for providing insight into the data and sharing useful information about the business of crowdsourcing.

1 Introduction

Contests have a rich history as a mechanism for the procurement of innovation in design and technology. With the growth of the internet, firms have begun using contests to procure ideas for advertising, new products and marketing strategies. For example, when motorcycle manufacturer Harley-Davidson split with its ad agency of 31 years, it turned to the crowd to create its next generation of advertising (Klaassen 2011). With the help of a crowdsourcing firm, Harley organized an *ideation contest* - fans of the brand could submit short ad ideas for a chance to win a cash prize. The winning submissions motivated a series of popular Harley marketing campaigns. Contests carry many advantages over the traditional ad agency model of advertising procurement: brands can expect a large number of ideas at a relatively low cost; participants tend to be actual end users of the product; and contests build awareness by engaging consumers in conversation with the brand (Pickering 2011, Kirby 2013).

Harley is not alone in adopting the contest model of ideation. Government agencies and firms in the private sector across a variety of industries have implemented ideation contests. For example, Challenge.gov, a government operated ideation platform, solicits ideas from participants for projects organized by different federal agencies, such as DARPA and NASA. Innocentive, a popular platform for scientific innovation, hosts ideation contests for companies such as Ford, GlaxoSmithKline and MasterCard. The crowdsourcing studio Tongal organizes advertising ideation contests for AT&T, Deloitte, General Electric, LEGO and Proctor & Gamble, among others.[1]

The success of an ideation contest hinges on its design - the choice of how to structure prizes and contest entry regulations. In this research, I empirically examine the impact of three broadly applicable design decisions - how many prizes to award, how much money to award per prize, and how many submissions to accept per participant - on contest participation and idea quality outcomes, such as expected total and maximum submission quality.

Prior research has explored how incentives affect ideation from an agency theory perspective (Toubia 2006, Girotra et al. 2010), but few papers have empirically examined the use of contest mechanisms for the procurement of ideas. I develop and estimate a structural model of ideation

[1] Some of the earliest ideation contests in marketing date back to the 1950s and 1960s (Kirby 2013). Popular brands would organize contests through newspapers and specialized publications to obtain ideas for ads, commercial jingles and new product names from consumers.

contests to assess the impact of different design parameters on contest outcomes. The model captures both participant and sponsor decision processes. Participants choose whether or not to enter a contest and how many ideas to submit based on their expected returns and costs of effort. The sponsor then ranks submissions by quality and rewards the winners.

Participants may differ in their *abilities* and *costs*. Ability heterogeneity reflects the notion that idea quality may differ across participants. Ideation contests attract a wide array of entrants with different backgrounds and experiences. Certain participants may submit higher quality ideas than others. For example, we may expect a Harley veteran to generate higher quality ideas for a motorcycle ad than someone with limited riding experience. Cost heterogeneity allows for participants to differ in how easy or difficult it is for them to think of ideas for a particular contest. For example, individuals with more outside commitments may have less time to participate in online contests, increasing their costs of making submissions. I allow for abilities and costs to differ by participant and contest. Moreover, participants can select into contests based on an unobservable (to the researcher) component of costs.

I use data from crowdsourcing platform Tongal to estimate the model in two stages. First, data on sponsor rankings of winning submissions identify participant abilities as a function of observable participant characteristics, such as demographics, past contest experience, and browsing behavior. The first stage amounts to a random-coefficients rank-ordered logit model, which I estimate using hierarchical Bayesian methodology. Second, participant submission decisions identify the costs of ideation. I estimate the second stage as an empirical discrete game, where participants choose how many ideas to submit to a given contest to maximize their expected payoffs. I use moment inequalities to partially identify the average cost function. This methodology allows for multiple equilibria, a non-parametric cost unobservable, and yields estimates that are robust to different specifications of participant information sets.

Counterfactual simulations reveal the impact of alternative prize allocation and submission limit decisions on contest outcomes under different assumptions about the information sets of participants. I experiment with two information structures, which I label complete and incomplete information. In the complete information scenario, participants know their own characteristics, as well as sponsor preferences and the characteristics of their competitors. In the incomplete information scenario, participants do not know sponsor preferences or competitor characteristics, but are aware of the joint density of these variables conditional on contest structure. I find that both information structures imply similar counterfactual outcomes on average across contests. However, the outcome of each individual contest may differ depending on the informational assumption.

First, I investigate the impact of offering a single prize instead of multiple prizes. I find that although multiple prizes motivate lower ability participants and demotivate higher ability participants, the number of prizes, holding fixed total award, has a negligible impact on participation and quality - the change in expected marginal returns to most participants is small compared to submission costs. Second, I explore the impact of increasing prize money. I find that a strong response from higher ability participants leads to an increase in idea quality but may not lead to a substantial increase in the total number of entrants. Finally, I examine the effect of reducing the maximum number of submissions allowed per participant. This policy benefits lower ability participants who would have otherwise been discouraged from entry by the presence of higher ability participants who submit multiple times to the same contest. A more stringent submission limit restricts higher ability participants, increasing the number of entrants but reducing expected quality outcomes.

The remainder of this paper is organized as follows. Section 2 summarizes the relevant theoretical and empirical literature on contest design. Section 3 presents the data and descriptive evidence of the importance of prize allocation. Section 4 outlines the structural model and Section 5 details the two-stage estimation routine. Section 6 examines the impact of counterfactual contest designs and presents practical implications. Section 7 concludes.

2 Contest Design

A contest is a game in which players invest costly effort in an attempt to win a prize. Throughout, I refer to players who consider entering a contest as *participants*. Of all participants, those who enter the contest are referred to as *entrants*, and the rest, as *non-entrants*. The *sponsor* organizes the contest and ultimately selects winners and awards prizes. Effort in the contest literature is typically viewed as a non-negative continuous decision variable. I view effort as the discrete number of idea *submissions* a participant makes to a given contest.[2]

[2]The ideation contests I study require participants to submit 140 character ideas for ads. Each participant can submit at most 5 ideas to a single contest. Section 3.1 and Appendix A.2 show evidence that the number of submissions is a good measure of participant effort - submissions react in expected ways to changes in prize allocation.

Traditionally, contests have been modeled as either imperfectly discriminating (Tullock 1980), all-pay auctions (Baye et al. 1994), or rank-order tournaments (Lazear and Rosen 1981). Imperfectly discriminating contests and rank-order tournaments typically allow for uncertain outcomes - the participant exerting the highest effort is not guaranteed to win. However, a higher effort increases the participant's chances of winning. In all-pay auctions, highest effort typically guarantees victory. Ideation contests share similarities with imperfectly discriminating contests and rank-order tournaments - participants who submit the most ideas are not guaranteed to win, and in contests with multiple prizes, submissions are ranked in order of the sponsor's preferences.

A key aspect of ideation contests is participant heterogeneity. Participants, with different levels of skill and experience, can freely join the platform and enter contests. Although a greater number of entrants improves the sponsor's chances of obtaining an extreme-value, high quality idea, especially in contests with significant participant uncertainty about sponsor preferences (Boudreau et al. 2011), increased participant asymmetries typically result in reduced effort (Baye et al. 1993, Stein 2002). Intuitively, participants with a low chance of winning are discouraged and "give up," which in turn reduces the level of competition for participants with a high chance of winning, resulting in a lower level of effort from all types. However, an appropriate choice of prize allocation can mitigate this concern.[3]

Theory literature has examined the impact of prize allocation on the effort of heterogeneous participants. Moldovanu and Sela (2001) explore the impact of multiple prizes on effort in all-pay auctions. The authors show that, holding fixed total award, a greater number of prizes encourages weaker participants, as they have a chance of winning one of the lower ranking prizes. On the other hand, stronger participants exert less effort, as with multiple prizes, the payoff from "losing" increases. The optimality of offering multiple prizes depends on participant heterogeneity and the convexity of their costs of effort. If costs are sufficiently convex, a smaller number of prizes will not encourage stronger participants to increase effort by enough to compensate for the reduced effort of weaker participants, and the sponsor may find it optimal to offer multiple prizes. Szymanski and Valletti (2005) argue that stronger participants may increase effort in response to multiple prizes in imperfectly discriminating contests. The added uncertainty of winning may motivate

[3]Fullerton and McAfee (1999) argue that restricting entry can also benefit sponsors. By imposing an appropriate entry auction mechanism that encourages higher ability participants to enter, the sponsor can expect greater effort while minimizing the costs of procurement.

stronger participants to react to increasing competition from weaker participants. Few papers have examined the impact of prize allocation on outcomes other than effort. Terwiesch and Xu (2008) consider expected maximum and average quality outcomes in imperfectly discriminating innovation contests and all-pay auctions. The authors similarly show that multiple prizes may be optimal in contests with heterogeneous participants, but a single prize works best for contests with ex-ante identical participants.[4] Overall, the effect of multiple prizes on effort is ambiguous and depends on participant heterogeneity and cost function shape.[5] I contribute to the literature by presenting estimates of different prize allocation policies on participation and quality outcomes, and suggesting practical implications for ideation contest design.

Although the question of how many submissions to accept per participant is unique to contests where participants can make multiple submissions, researchers have investigated the related aspect of restricted bidding in all-pay auctions. Che and Gale (1998) consider the impact of caps on investments in political lobbying in an all-pay auction with one high-valuation (strong) player and one low-valuation (weak) player. The authors find that bid caps can increase total spending by limiting the strong participant and encouraging the weak participant. Che and Gale (2003) similarly show that handicapping a stronger participant in research contests can improve the contest outcome. I investigate the impact of restricting the number of submissions per participant - a relevant and easy to implement policy in the context of ideation contests. My results show that submission limits constrain higher ability participants and increase overall entry. However, I find that a more stringent submission limit may reduce expected total and maximum idea quality.

Substantial progress in the empirical literature on contests has been achieved with the increasing availability of online data. Boudreau et al. (2015) examine the impact of competition on the effort of heterogeneous participants in the context of the popular TopCoder platform for programming contests. The authors examine a number of contest design policies but do not focus on the question of how many prizes to award or how many submissions to accept per participant. Furthermore, the authors adopt the all-pay auction model of contests (Moldovanu and Sela 2001), whereas I focus on

[4]See Sisak (2009) for a survey of the theoretical literature on multiple prizes in contests.

[5]Apart from heterogeneity, participant risk-aversion may also motivate the adoption of multiple prizes. Kalra and Shi (2001) show that in rank-order sales tournaments with sufficiently risk-averse homogeneous participants, multiple prizes may increase effort. However, experimental research suggests that risk-averse participants are less likely to enter tournaments altogether (Eriksson et al. 2009, Dohmen and Falk 2011). Throughout, I focus on contests with risk-neutral, heterogeneous participants.

the imperfectly discriminating framework (Stein 2002, Szymanski and Valletti 2005), which I find more relevant for ideation contests where the outcome depends heavily on idiosyncrasies in sponsor preferences. Yoganarasimhan (2015) presents a model of beauty contest auctions, or procurement auctions with uncertain outcomes. The author estimates the model using data from a freelance marketplace. In beauty contest auctions, participants bid a dollar fee, and the winner receives her fee as compensation from the sponsor. Ideation contests differ in that participants invest costly effort to think of ideas, and the winners receive pre-announced prizes. In the setting of online design contests, research has explored the impact of feedback and entry visibility on participation and submission quality (Wooten and Ulrich 2011, 2013) as well as the effects of competition on experimentation (Gross 2014). I present an estimable model of ideation contests with multiple prizes, participant heterogeneity, and the possibility of selection into contests based on costs of effort. I contribute to the empirical literature by presenting estimates of the impact of a number of key design parameters, such as the number of prizes, prize amount, and submission limit, on participation and idea quality outcomes. Furthermore, I address a recent call in literature to allow for more flexibility in the information structures of empirical games (Borkovsky et al. 2015) and derive contest outcome predictions that are robust to different informational assumptions.

3 Data

I use data from Tongal, a popular crowdsourcing platform. Tongal organizes contests for brands across a variety of product categories, including consumer packaged goods, body care and health products, electronics, pet supplies, and toys.

Ideation contests on the Tongal platform operate as follows. Tongal and the contest sponsor jointly decide on how many prizes to offer and how much money to offer per prize. The sponsor presents participants with the contest prize allocation, rules and regulations, and a description of the ideation topic. Participants can then enter the contest by submitting at least one 140 character idea for an ad based on the topic suggested by the sponsor. Each entrant can submit at most 5 ideas to a single contest. After the contest ends, the sponsor reviews the content of each submission without knowledge of the identity or characteristics of its creator. Winning submissions are ranked and their creators receive prize money. The platform does not display the identities or actions of participants during the contest period. Only after the sponsor selects winners does the platform make public the list of winning submissions.

I focus on a sample of 181 ideation contests that ran from 2011 to 2015 (the platform was founded in 2009) and a set of 8,875 participants who entered at least one of these contests. For each contest, I observe the number of submissions made by each entrant, the number of prizes awarded and total prize amount. All contests divide prizes evenly among winners. For example, each winning submission receives $250 if a contest offers 4 prizes with a total award of $1,000.

An important aspect of many contests is that not all participants who consider entering choose to do so. I use browsing data to define the set of likely non-entrants, or participants who considered entering a contest but chose not to. Specifically, participants who did not enter the contest but viewed the contest page more than once and were active in the past 3 months are considered likely non-entrants. I restrict non-entrants to this subset to avoid including participants who were simply "surfing" the site without seriously considering entry into the contest. This procedure yields a total of 9,732 instances of non-entry by likely non-entrants. On 35,011 occasions, participants make at least one submission.

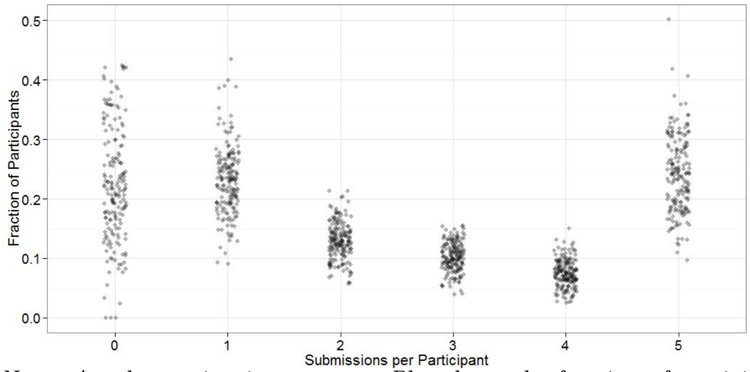
Table 1: Summary of Contest Characteristics

Per-Contest Characteristics	Min	Median	Mean	Max
Non-Entrants	0	48	54	124
Entrants	58	187	193	499
Submissions	178	551	572	1,875
Number of Prizes	1	4	5	50
Prize Amount per Spot	$100	$250	$323	$1,250
Total Award	$500	$1,000	$1,450	$10,000

Table 1 presents summary statistics for the contests considered. The contests tend to attract a high number of entrants and submissions, with the average contest securing 193 entrants and 572 submissions. There is also substantial variation in prize allocation across contests, with the number of prizes ranging from 1 to 50 and prize amount per winning spot ranging from $100 to $1,250. Figure 1 presents a plot of the distribution of submissions per participant with a contest for all 181 contests in the data. A significant proportion of participants does not enter, submits once, or makes the maximum number of submissions allowed.

For each one of the 8,875 participants, in addition to browsing data, I observe a set of characteristics collected by the platform. These include: demographic variables such as age, gender, and

Figure 1: Distribution of Submissions within Contests



Note: An observation is a contest. Plot shows the fraction of participants who made d submissions within each contest, where $d \in \{0, ..., 5\}$.

country; platform-user characteristics, such as registration date and whether or not the user has a profile picture; and past performance variables, such as past victories and earnings. Appendix A.1 presents summary statistics for the observed participant characteristics and draws comparisons between the characteristics of participants who won and participants who did not win.

3.1 Descriptive Evidence

Is there evidence in the data that participants change their actions in response to different prize allocations? Such evidence would suggest that prize allocation is an important design parameter that can alter behavior. I run fixed effects regressions to investigate the impact of the number of prizes and prize amount on the number of submissions made by participants.

First, consider the impact of the number of prizes on submissions. Contests offering the same total award are grouped into a single category. I compare the number of submissions made by the same participant across contests with a different number of prizes within the same category. The regression equation is given by

$$Y_{it} = \alpha X_t + \gamma Z_{it} + \xi_{iC_t} + \epsilon_{it}, \tag{1}$$

where Y_{it} is the number of submissions made by participant i in contest t, $X_t = 1$ if a contest offers a larger number of prizes than the average within its category, and $X_t = 0$ otherwise. The Z_{it} term and the associated parameter vector γ capture the impact of contest-varying participant characteristics on submissions. In particular, Z_{it} includes the participant's page views and total earnings at the start of the contest. Participant-category specific fixed effects ξ_{iC_t} control for unobserved participant heterogeneity, where C_t denotes the category of contest t. Finally, α is the parameter of interest and ϵ_{it} is an error term.

I label participants as either *heavy* or *light* users. Heavy users enter in more than three contests in the data. Light users enter at most three contests. I expect heavy users to have lower costs or higher abilities and a greater chance of winning than light users. The extent to which heavy and light users behave differently will shed light on the importance of participant heterogeneity in assessing the impact of different prize allocation decisions.

Table 2 presents the parameter estimates for Regression 1, estimated separately for heavy and light users. The results suggest that heavy users submit less to contests with a greater number of smaller prizes. Although I do not find a significant effect for light users, the regression estimates for heavy users are consistent with the predictions of theoretical models such as Moldovanu and Sela (2001). Higher ability entrants may be discouraged by multiple prizes.

Table 2: Regression of Submissions on Number of Prizes

	Heavy Users	Light Users
More Prizes than Category Average	-0.154***	0.080
	(0.037)	(0.139)
Participant-Category Fixed Effects	Y	Y
Contest-Varying Participant Characteristics	Y	Y
R^2	0.089	0.166
Observations	35,763	8,980

*Note: An observation is a contest-participant pair. "More Prizes than Category Average" is an indicator for whether or not the contest offers more prizes than the average for its category C_t. All contests within the same category offer the same total award. Contest-varying participant characteristics include the participant's page views and total earnings at the start of the contest. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Although Regression 1 allows for unobserved heterogeneity across participants and across categories for the same participant, it does not capture sources of unobserved heterogeneity correlated with X_t across contests within the same category. For example, contests that offer multiple prizes may require more creativity and have a higher difficulty level, resulting in a reduced number of submissions from heavy users. However, I focus only on 140 character ad ideation contests, which require completion of substantially similar tasks from participants, mitigating concerns about differences in difficulty across contests.

Next, I consider the impact of prize amount on submissions. Regression 1 can be modified as follows. Group all contests offering the same number of prizes into the same category and let C_t denote the category of contest t. Let X_t equal the total award for contest t. Now, the identifying variation comes from differences in submission decisions made by the same participant for contests offering the same number of prizes but different prize amounts. Table 3 summarizes the regression results. Increasing prize amount appears to increase submissions from heavy users, but no significant increase is found for light users, consistent with the predictions of theoretical contest models with heterogeneous participants, such as Stein (2002). Higher ability entrants have a greater chance of winning a prize, and hence, may be more heavily influenced by changes in prize amount.

Table 3: Regression of Submissions on Prize Amount

	Heavy Users	Light Users
Prize Amount	0.171***	0.089
	(0.018)	(0.069)
Participant-Category Fixed Effects	Y	Y
Contest-Varying Participant Characteristics	Y	Y
R^2	0.091	0.126
Observations	35,763	8,980

*Note: An observation is a contest-participant pair. "Prize Amount" refers to the total prize money awarded in thousands of dollars. All contests within the same category offer the same number of prizes. Contest-varying participant characteristics include the participant's page views and total earnings at the start of the contest. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

The descriptive evidence presented in this section suggests that submission decisions respond to changes in prize allocation. In Appendix A.2, I present further evidence that the results are not sensitive to different regression specifications and different characterizations of heavy and light users.

To assess the impact of contest design on quality, a variable not observed in the data, I require a structural model of sponsor preferences. Furthermore, a model of participant submission decisions would enable an analysis of the impact of submission limits, a variable that remains unchanged across contests in the data (all contests allow for at most 5 submissions per entrant). I proceed to suggest a structural model that would enable the derivation of precise predictions as to how different types of participants react to changes in prize allocation and entry restrictions.

4 Model

I model each ideation contest as an independent game consisting of two stages. First, participants considering entry decide on how many submissions to make given their costs and expected payoffs. I consider a trial-and-error model of ideation, whereby participants sample from a quality distribution with each submission in an attempt to generate an idea of high quality for the sponsor. This approach is common in models based on the statistical view of innovation in new product design (Dahan and Mendelson 2001, Loch et al. 2001) and in models of research contests with uncertain outcomes (Taylor 1995, Fullerton and McAfee 1999, Terwiesch and Xu 2008). Second, the sponsor reviews all submissions and ranks the top submissions by quality. Most sponsors in the data offer only one contest. Sponsors who offer multiple contests tend to focus on different products and ideation topics across contests. I use sponsor and contest interchangeably and denote both by t.

The model will capture three key features of ideation contests. First, participant abilities may differ across contests. If two participants make the same number of submissions, the participant with the higher ability has a larger expected payoff. Ability heterogeneity captures the notion that the "fit" between a participant and a contest may depend on the participant's background and the contest ideation topic. Second, participants exhibit cost heterogeneity and may select into contests they find most convenient based on an unobservable component of costs. If two participants have the same ability but different costs, the participant with the lower cost may increase her expected payoff by making a larger number of submissions. Cost heterogeneity accounts for the possibility that certain participants may be busier than others at different times, or find it more difficult to think of ideas for certain contests. Third, participants may view their own abilities, the number of competitors, competitor abilities and competitor actions with uncertainty and form expectations of their own expected payoffs given their information sets.

I work backwards and first present the model for the second stage sponsor decision (Section 4.1), followed by the model for participant entry decisions (Section 4.2). The empirical implementation of the two stages is presented in Section 5.

4.1 Sponsor Choice Model

Consider the sponsor's decision process after it receives a set of submissions. From the perspective of the sponsor, submission s by participant i in contest t has quality

$$q_{st} = \beta_t X_{it} + \epsilon_{st},$$

where X_{it} is a vector of participant characteristics, β_t is a vector of the sponsor's preferences, and ϵ_{st} is a mean-zero quality shock to submission s. Let $a_{it} = \beta_t X_{it}$ denote participant i's ability in contest t. Participants with a higher ability will tend to make higher quality submissions than participants with a lower ability. Moreover, if a participant has a high ability in one contest, she may not necessarily have a high ability in a different contest. For example, a participant based in North America may submit higher quality ideas for ads targeted at North American audiences but may not perform as well in contests for European brands. The interaction of heterogeneous sponsor preferences β_t and participant characteristics X_{it} reflects the differences in participant submission quality across contests.

The shock ϵ_{st} captures all heterogeneity in submission quality that cannot be explained by ability. Participant i draws a separate shock for each submission. The average quality of a participant's submissions is simply a_{it}, but each submission may be of higher or lower quality than the average depending on the participant's draws of quality shocks. I assume that the participant knows the distribution of ϵ_{st} before she chooses to submit but never learns of the realizations of her ϵ_{st} draws. As a result, the participant may choose how much to submit based on her ability a_{it} but must form expectations of her payoffs with respect to the distribution of quality shocks.

The sponsor observes q_{st} for each s and ranks submissions by quality. Only the best N_t submissions receive a ranking, where N_t is at least as large as the number of prizes.[6] In other words, the sponsor chooses a ranking $s_{(1)}, ..., s_{(N_t)}$ such that $q_{s_{(1)}t} \geq q_{s_{(2)}t} \geq ... \geq q_{s_{(N_t)}t} \geq q_{kt}$, where q_{kt} is the quality of any other submission k not in $s_{(1)}, ..., s_{(N_t)}$.

[6]The number of submissions that recieve a ranking varies across contests but is usually equal to the number of prizes.

4.2 Participant Entry Model

Risk-neutral participants form expectations of their contest payoffs with respect of the distribution of quality shocks and their perceptions of competitor actions and abilities. Suppose that a total of I_t participants consider entering contest t. Participant i with ability $a_{it} = \beta_t X_{it}$ chooses to make $d_{it} \in \{0, 1, ..., D\}$ submissions in contest t, where D is the submission limit. Expected payoffs are given by

$$\pi_{it} = E\left[R_t(d_{it}, d_{-it}; a_{it}, a_{-it})|\mathcal{J}_{it}\right] - c_{it}(d_{it}).$$

The expected returns function $R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$ captures the expected winnings of a participant with ability a_{it} who makes d_{it} submissions, given competitor abilities a_{-it} and actions d_{-it}.[7] The expected returns function depends on participant and competitor characteristics, sponsor preferences, the distribution of quality shocks, the number of prizes, and the prize amount offered by the sponsor. Participants know their own X_{it} and the contest prize structure but may view sponsor preferences, the number of competitors, competitor characteristics and competitor actions as random variables because of incomplete information. A participant forms an expectation of her expected returns function given her information set $\mathcal{J}_{it}$.

I specify the cost function as

$$c_{it}(d_{it}) = (\theta_1 + \nu_{it})d_{it} + \theta_2 d_{it}^2,$$

where ν_{it} is a mean-zero participant-contest specific cost unobservable, and θ_1 and θ_2 are cost parameters. Prior to entry, each participant observes her cost shock ν_{it} and chooses how many submissions to make to maximize expected payoffs π_{it}. She may also choose to make no submissions and obtain zero payoffs.

5 Estimation

Estimation proceeds in two stages. I first estimate the sponsor choice model using hierarchical Bayesian methodology. Given the first stage results, I estimate the participant entry model using

[7] For example, in a contest with one prize, the expected winnings of a participant are simply equal to the prize amount, multiplied by the probability that the participant generates at least one submission of higher quality than all other submissions, given the participant's ability and action, and competitor abilities and actions.

moment inequalities. The underlying game is likely to have multiple equilibria because of the discrete action space. The moment inequalities methodology allows for multiple equilibria, does not require explicit specification of participant information sets, and permits a flexible distribution of cost unobservables. In the second stage, I follow the estimation procedure for discrete games with ordered choices suggested by Ishii (2008) and Pakes et al. (2015).

5.1 First Stage: Sponsor Choice Model

I use data on sponsor ranking decisions and participant characteristics to estimate the sponsor choice model. To proceed, I must specify the distribution of quality shocks.

Assumption 1 *The independent and identically distributed quality shocks ϵ_{st} follow a demeaned type 1 extreme value distribution and are not known to participants before they make submission decisions.*

With each submission, the participant makes an independent draw from an extreme value quality shock distribution. The average quality of submissions made by the participant is given by the participant's ability. Assumption 1 ensures that participants cannot select into contests based on a component of ability unobserved to the researcher and reduces the complexity of the estimation routine.[8] Although participants cannot select into contests based on unobservable components of their abilities, the participant entry model allows for selection based on cost unobservables.

Next, I require an assumption on the distribution of sponsor preferences. Across all contests, sponsors rank a total of 905 submissions, whereas for each contest, I observe on average 5 ranking decisions (equal to the average number of prizes per contest). Absent a distributional assumption, it is not possible to recover sponsor-specific preferences with sufficient precision.

Assumption 2 *Sponsor preferences β_t follow a multivariate normal distribution with mean μ and covariance matrix Σ.*

This assumption implies that estimates of β_t will depend on the sample of contests used in estimation, and furthermore, that different types of sponsors do not strategically select to post a

[8] To test the impact of this assumption, I estimated the model and re-ran counterfactuals after excluding the most significant observable component of participant abilities. Estimates changed, but the counterfactual implications remained qualitatively similar. Furthermore, in Section 5.1.4, I illustrate that observable participant characteristics can explain a significant portion of the variance in contest winners.

contest on the platform. It is straightforward to augment the model to allow for sponsor preferences to depend on contest characteristics.

5.1.1 Identification

Identification relies on rankings data and heterogeneity in participant characteristics. Sponsors face submissions from participants with different characteristics and choose rankings based on these choice sets. Variation in sponsor decisions for different sets of entrant characteristics identifies components of the mean population preference vector μ. Rankings data yield multiple decision instances per sponsor. For example, if the sponsor ranks two submissions, each ranking can be viewed as an independent choice. The sponsor assigns first rank to the highest quality submission out of the set of all submissions. Then, the sponsor assigns second rank to the highest quality submission out of the set of all submissions excluding the first-ranked submission. Identification of the covariance of sponsor-specific preferences Σ is aided by the availability of multiple decision instances per sponsor.

5.1.2 Likelihood

The likelihood of observing a ranking $s_{(1)}, ..., s_{(N_t)}$ is

$$L_t\left(s_{(1)}, ..., s_{(N_t)}|\beta_t\right) = \prod_{r=1}^{N_t}\left(\frac{\exp\{\beta_t X_{s_{(r)}t}\}}{\sum\limits_{j=r}^{N_t}\exp\{\beta_t X_{s_{(j)}t}\} + \sum\limits_{k\in\emptyset}\exp\{\beta_t X_{kt}\}}\right),$$

where N_t is the number of ranked submissions, $\emptyset$ is the set of all unranked submissions, and $X_{st} = X_{it}$ if submission s belongs to participant i. Integrating over β_t yields the likelihood of the data, corresponding to the likelihood of a random-coefficients rank-ordered logit model:

$$\mathcal{L} = \int \prod_{t=1}^{T} L_t\left(s_{(1)}, ..., s_{(N_t)}|\beta_t\right) f(\beta) d\beta.$$

Prior research has used rank-ordered logit models (also known as exploded logits) to recover preferences from rankings in consumer survey data (Beggs et al. 1981, Chapman and Staelin 1982). In my setting, a structural model of sponsor choice generates a statistical rank-ordered logit model that can be estimated with random coefficients using data on sponsor rankings of contest winners.

I use hierarchical Bayesian methods to estimate the model. Posterior densities of the parameters are obtained using Markov-Chain Monte Carlo (MCMC) integration. In my application, the hierarchical Bayesian estimation technique has an advantage over classical alternatives, such as simulated maximum likelihood (SML). Namely, the MCMC procedure recovers sponsor-level parameters β_t as a byproduct of the estimation routine, whereas in SML, individual sponsor preferences are treated as nuisance parameters and integrated out (Rossi et al. 2012). I require sponsor-specific preferences for the participant entry model to obtain an expression for $R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$ for each participant i and contest t. As a result, I choose to employ a hierarchical Bayesian estimation methodology in the first stage.

5.1.3 Participant Characteristics

I use a set of characteristics collected by the platform to account for possible sources of ability heterogeneity. Variables in the set of characteristics X_{it} include demographics, participant-platform characteristics, and measures of skill and interest in the contest. Demographics consist of indicators for gender, country, race, relationship status and education, and a continuous age variable. Participant-platform characteristics include indicators for adoption date, light user status, whether or not the participant was referred to the platform, and whether or not the participant had a profile picture. Skill and interest is summarized by indicators for whether or not the participant previously entered a contest or joined during the current contest, producer status, and functions of page views and past earnings. These characteristics should be interpreted not as variables that have a causal effect on participant abilities but rather as variables that proxy for sources of heterogeneity in participant ability that is invariant to contest design. Otherwise, contests may no longer be treated as independent games. Table 4 presents the definitions for all variables used in estimation.

5.1.4 Ability Estimates

The leftmost columns of Table 5 present parameter estimates from the heterogeneous sponsor choice model. Not surprisingly, heavy users perform better in contests and have higher abilities than light users. Participants who view the contest page more and who have higher past earnings also have higher abilities. Producers, or participants with film-making experience, perform better

Table 4: Description of Participant Characteristics

Variable	Definition
Demographics	
Gender_i	1 participant i is female and 0 otherwise.
Country_i	1 if participant i is from the US and 0 otherwise.
Race_i	1 if participant i is listed as Caucasian and 0 otherwise.
Relationship_i	1 if participant i is in a relationship and 0 otherwise.
Education_i	1 if participant i has at least a college degree and 0 otherwise.
Date_of_Birth_i	Standardized date of birth of participant i.
Participant-Platform Characteristics	
Early_Adopter_i	1 if participant i joined before the end of 2011 and 0 otherwise.
Late_Adopter_i	1 if participant i joined after the start of 2013 and 0 otherwise.
Light_User_i	1 if participant i entered no more than 3 contests and 0 otherwise.
Referred_i	1 if participant i was referred to the platform and 0 otherwise.
Picture_i	1 if participant i has a profile picture and 0 otherwise.
Skill and Interest	
$\text{First\_Contest}_{it}$	1 if t is participant i's first contest and 0 otherwise.
$\text{Joined\_During}_{it}$	1 if participant i joined during contest t and 0 otherwise.
Producer_i	1 if participant i has video production skills and 0 otherwise.
$\text{Past\_Earnings}_{it}$	Logarithm of 1 plus participant i's winnings up to contest t.
Page_Views_{it}	Logarithm of 1 plus participant i's page views of contest t.

in ideation contests. A profile picture is also a strong sign that a participant has a high ability. A participant with a profile picture may be more invested in the platform than a participant with no profile picture. Interestingly, participants referred to the platform perform worse, perhaps because participants who join on their own initiative have a greater interest in ideation. The variance column of Table 5 shows evidence of significant heterogeneity in sponsor preferences across contests. The MCMC method yields estimates for sponsor preferences at the level of each contest. I recover the posterior means for each sponsor and obtain estimates of participant abilities in each contest. Overall, the estimates of sponsor preferences point to a high level of heterogeneity in participant abilities, both within and across contests.

For comparison, the rightmost column of Table 5 lists the estimates obtained from a model with no unobserved heterogeneity in sponsor preferences. I investigate the explanatory power of each model by looking at how many winners are excluded from consideration if I rank participants in each contest by their ability (based on each model) and exclude a fraction of the lowest ranking participants. The higher the explanatory power of the model, the lower the number of actual winners that are filtered out. Figure 2 shows the results. Interestingly, the abilities implied by the heterogeneous model identify winners very accurately. If 50% of all participants are excluded based on their abilities, the set of remaining participants includes over 90% of the actual winners. If 90%

Table 5: Posterior Means of Sponsor Choice Model Population-Level Parameters

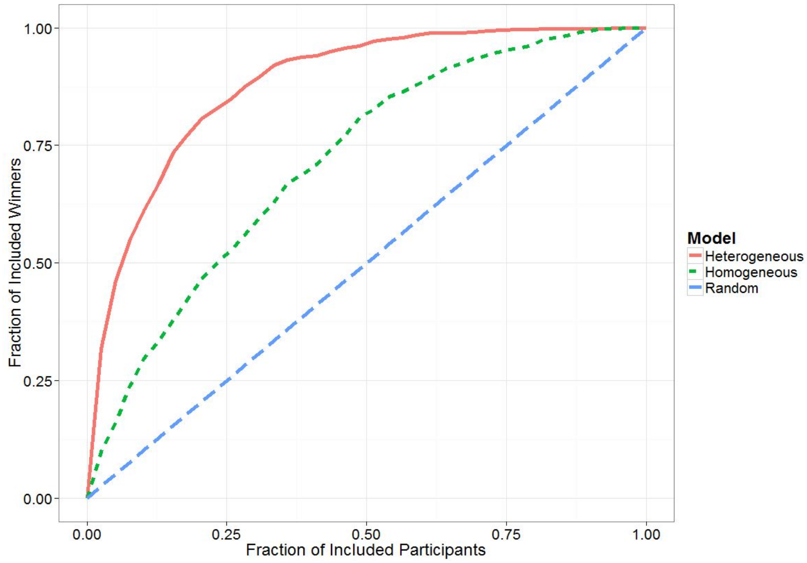
Variable	Mean (μ)	Variance (diagonal of Σ)	Homogeneous Model
Demographics			
Gender	-0.174	0.407	-0.010
	(0.058)	(0.082)	(0.080)
Country	0.303	0.496	0.129
	(0.077)	(0.099)	(0.099)
Race	-0.164	0.541	-0.190
	(0.082)	(0.102)	(0.104)
Relationship	-0.099	0.552	0.176
	(0.074)	(0.103)	(0.103)
Education	-0.162	0.422	-0.024
	(0.057)	(0.070)	(0.084)
Date of Birth	0.187	0.410	0.117
	(0.055)	(0.060)	(0.038)
Participant-Platform Characteristics			
Early Adopter	-0.330	0.787	-0.017
	(0.103)	(0.156)	(0.119)
Late Adopter	0.306	0.550	0.322
	(0.077)	(0.114)	(0.093)
Light User	-1.313	1.130	-1.288
	(0.127)	(0.234)	(0.183)
Referred	-0.384	0.454	-0.314
	(0.062)	(0.090)	(0.096)
Picture	0.190	0.529	0.045
	(0.069)	(0.095)	(0.099)
Skill and Interest			
First Contest	0.099	0.673	0.419
	(0.119)	(0.103)	(0.132)
Joined During	-0.069	0.893	0.208
	(0.143)	(0.195)	(0.161)
Producer	0.207	0.415	0.337
	(0.054)	(0.055)	(0.087)
Past Earnings	0.169	0.108	0.087
	(0.017)	(0.016)	(0.011)
Page Views	0.491	0.597	0.379
	(0.067)	(0.083)	(0.052)

Note: Standard deviations in parentheses. "Homogeneous Model" column shows estimates obtained from a rank-ordered logit model with homogeneous coefficients.

of all participants are excluded (which amounts to an elimination of roughly 80,000 submissions), the set of remaining participants includes over half of all winners. The homogeneous model does not perform as well, but does outperform the baseline scheme of randomly selecting and excluding participants. These results suggest that observable participant characteristics do well in explaining which submissions win or lose in my sample of contests, and moreover, that unobserved sponsor preference heterogeneity plays an important part in increasing the explanatory power of the model.

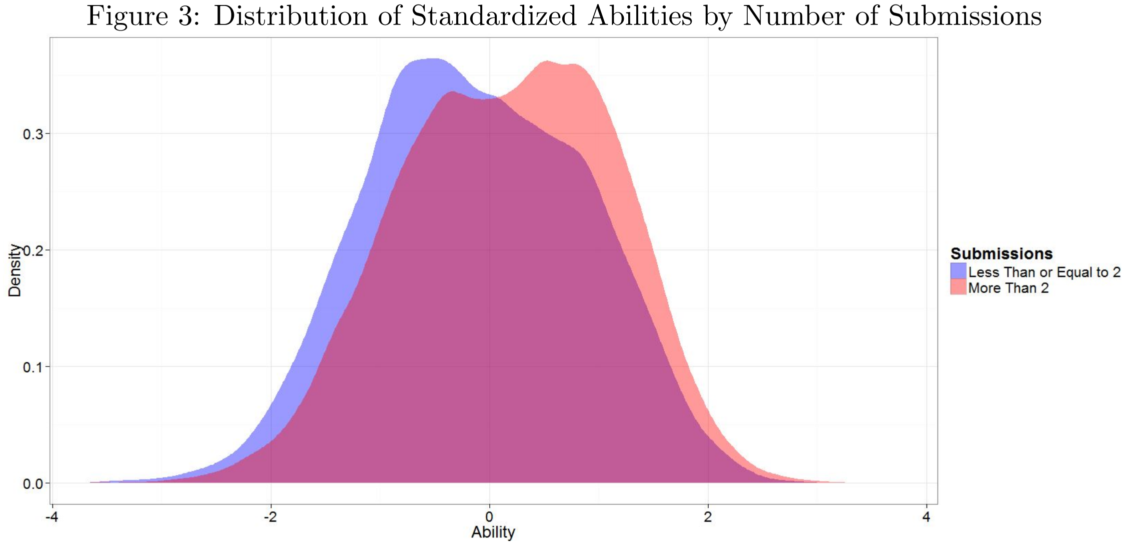
To test whether the estimated abilities are reasonable, I plot densities of abilities for participants who submitted more than twice and participants who submitted no more than twice, standardized at the level of each contest. Figure 3 shows that participants who submit more than twice in a single contest tend to have higher abilities than participants who submit no more than twice in

Figure 2: Explanatory Power of Sponsor Choice Model



the same contest, consistent with theoretical predictions. A Kolmogorov-Smirnov test rejects the null hypothesis that the density of abilities for participants who submit less does not stochastically dominate the density of abilities for participants who submit more with a p-value less than 0.01.

Figure 3: Distribution of Standardized Abilities by Number of Submissions



Note: Abilities standardized within contests.

5.2 Second Stage: Participant Entry Model

I use moment inequalities to partially identify cost parameters θ_1 and θ_2. Pakes et al. (2015) show how moment inequalities can be used to obtain upper and lower bounds on cost parameters for discrete choice games where agents make ordered choices. With moment inequalities, I need not explicitly specify an equilibrium selection mechanism. Furthermore, the methodology allows for a flexible distribution of cost unobservables and yields estimates that are robust to different specifications of participant information sets. However, parameters will typically be set identified and not point identified. In other words, moment inequalities yield a set of parameters as opposed to a point, and confidence bounds must be obtained taking this into account.

The participant entry model can be rewritten as follows. I define the expectational error $\omega_{itd_{it}}$ as the difference between a participant's expected and actual returns based on the sponsor choice model: $\omega_{itd_{it}} = E\left[R_t(d_{it}, d_{-it}; a_{it}, a_{-it})|\mathcal{J}_{it}\right] - R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$. Sources of expectational error may include participant uncertainty about competitor actions (as a function of costs) and abilities, and may also incorporate optimization mistakes made by the participant in evaluating her expected returns. Then, the payoff equation can be written as

$$\pi_{it} = R_t(d_{it}, d_{-it}; a_{it}, a_{-it}) - c_{it}(d_{it}) + \omega_{itd_{it}}.$$

I require that the expectational errors $\omega_{itd_{it}}$ are mean-zero conditional on the participant's information set and, at this stage, place no additional restrictions on their distribution.

I proceed by first deriving a lower bound for marginal costs, where I take into account the possibility that participants who made the maximum number of submissions may have had particularly low costs. Then, I derive an upper bound for marginal costs, where I use a selection correction technique to account for the possibility that non-entrants may have had particularly large costs. Additional assumptions about the distributions of $\omega_{itd_{it}}$ and ν_{it} are introduced as they become relevant.

5.2.1 Lower Bound

First, consider the derivation of the lower bound for marginal costs. Define a function of the difference in observable returns from making one additional submission as

$$\Delta R^*_{it}(d_{it}+1, d_{it}) = \begin{cases} R_t(d_{it}+1, d_{-it}, a_{it}, a_{-it}) - R_t(d_{it}, d_{-it}, a_{it}, a_{-it}), \text{ if } d_{it} < 5, \\ 0, \text{ if } d_{it} = 5, \end{cases}$$

and let $\omega_{itd_{it}+1,d_{it}} = \omega_{itd_{it}+1} - \omega_{itd_{it}}$. By revealed preference, for a participant who made less than 5 submissions,

$$\underbrace{\Delta R^*_{it}(d_{it}+1, d_{it}) + \omega_{itd_{it}+1,d_{it}}}_{\text{expected marginal return}} \leq \underbrace{\theta_1 + \theta_2(2d_{it}+1) + \nu_{it}}_{\text{marginal cost}},$$

as the expected marginal return from making one additional submission must be no greater than the marginal cost of making one additional submission. Otherwise, the participant would have made $d_{it}+1$ instead of d_{it} submissions. For a participant who made 5 submissions, the expected marginal return from making one additional submission is likely an overestimate of the marginal cost of doing so, as the participant may have chosen to make more submissions under a more lenient submission limit. I make the assumption that the marginal cost of making one additional submission is at least zero for entrants who made the maximum permitted number of submissions.

Assumption 3 *The condition $\theta_1 + \theta_2(2d_{it}+1) + \nu_{it} \geq 0$ holds for entrants with $d_{it} = 5$.*

Let z^k_{it} be a positive variable, known to the participant at the time of submission and orthogonal to the participant's cost unobservable, such that $E[\omega_{itd_{it}} \mid z^k_{it}] = E[\nu_{it} \mid z^k_{it}] = 0$. I refer to z^k_{it} as the value of instrument k for participant i in contest t. Any participant or contest characteristic in the participant's information set that is not correlated with the participant's cost unobservable will qualify as an instrument. Taking the expectation over participants, it must be the case that

$$E\left[\underbrace{\theta_1 + \theta_2(2d_{it}+1)}_{\text{marginal cost}} - \underbrace{\Delta R^*_{it}(d_{it}+1, d_{it})}_{\text{marginal return}} \mid z^k_{it}\right] \geq 0.$$

Conditional on z^k_{it}, the expectational errors $\omega_{itd_{it}+1,d_{it}}$ average out to zero because z^k_{it} is an element of the participant's information set and participants are correct on average. The cost unobservables ν_{it} average out to zero because the expectation does not condition on the participant's action. The ability to take an expectation over cost unobservables for all participants, regardless of their action, is crucial for the unbiased (or at least conservative) estimation of bounds on cost parameters.

An empirical analogue for the lower bound for marginal costs can be written as

$$m^L_k(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\sqrt{I_t}}{I_t}\sum_{i=1}^{I_t} g(z^k_{it})\Delta r^*_{it}(d_{it}+1, d_{it};\theta),$$

where T is the total number of contests used in estimation, $g(z_{it}^k)$ is any positive-valued function of the variable z_{it}^k, the vector $\theta = (\theta_1, \theta_2)$, and

$$\Delta r_{it}^*(d_{it}+1, d_{it}; \theta) = \Delta R_{it}^*(d_{it}+1, d_{it}) - \theta_1 - \theta_2(2d_{it}+1).$$

I weight contest averages by the square-root of the total number of participants, as the averages for larger contests may exhibit less noise. Any θ that satisfies $m_k^L(\theta) \geq 0$ must lie in the identified set of cost parameters.

In practice, $R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$ is not analytically tractable in contests with many prizes but is required as an input to $\Delta r_{it}^*(d_{it}+1, d_{it})$ in the definition of $m_k^L(\theta)$. To obtain expected returns, it is necessary to consider the probability of observing all possible combinations of winning submissions from the set of all submissions. For contests with multiple prizes and hundreds of submissions, this expression can be analytically intractable. I use simulation to obtain the an approximation of the expected returns function for each participant in every contest.[9] Furthermore, participant abilities a_{it} must be approximated with the ability estimates based on the posterior means of sponsor preferences obtained through the sponsor choice model. Given the limited data available for each individual sponsor, it is not possible to obtain consistent estimates of participant-contest specific abilities. However, given that sponsors differ significantly in their preferences, ability estimates obtained from a random coefficients model will perform better than ability estimates obtained from a homogeneous model.[10]

5.2.2 Upper Bound

Next, consider the upper bound for marginal costs. For entrants i in $L_t = \{i : d_{it} > 0\}$, define the difference in observable returns from making one less submission as

$$\Delta R_{it}(d_{it}, d_{it}-1) = R_t(d_{it}, d_{-it}, a_{it}, a_{-it}) - R_t(d_{it}-1, d_{-it}, a_{it}, a_{-it}).$$

[9]It can be shown that simulation error averages out in the moment inequalities framework.

[10]To test the impact of using possibly biased ability estimates in $R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$, I estimated the cost parameters using abilities implied by both the homogeneous and the heterogeneous sponsor choice models. I found that despite the significant inherent bias, the homogeneous model abilities generated a very similar but slightly larger identified set.

Then, by revealed preference, for $i \in L_t$,

$$\underbrace{\Delta R_{it}(d_{it}, d_{it}-1) + \omega_{itd_{it},d_{it}-1}}_{\text{expected marginal return}} \geq \underbrace{\theta_1 + \theta_2(2d_{it}-1) + \nu_{it}}_{\text{marginal cost}}.$$

In other words, the expected marginal return of increasing submissions from $d_{it} - 1$ to d_{it} must have been greater than the associated marginal cost. Otherwise, entrants would have made one less submission than they actually did.

The above condition holds only for participants who submitted at least once. I must take into account the possibility that non-entrants, or participants with $d_{it} = 0$, may have had particularly large cost unobservables. If an empirical analogue, only for entrants, is developed based on the above inequality, the estimated upper bound on costs may be too low. Pakes et al. (2015) suggest using symmetry of the ν_{it} distribution to obtain an upper bound on the ν_{it} for non-entrants. Intuitively, the negative of the lowest lower bound for ν_{it} can be used as the highest upper bound for the negative of the ν_{it} of non-entrants. This result also holds as long as the ν_{it} density is not skewed left.

Assumption 4 *For each contest, the cost unobservables ν_{it} follow a mean-zero distribution that is not skewed left.*

For exposition, I derive all subsequent inequalities assuming that the ν_{it} follow a symmetric distribution, which will yield conservative inequalities if the actual distribution is skewed right. Assumption 4 allows for the cost unobservables to correlate with participant characteristics but requires that contests do not differ in difficulty level. In my analysis, I only consider similar contests for 140 character ideas. Furthermore, the relative dispersion of submissions is 0.84 on average within contests and 0.82 pooled across all contests, suggesting the absence of substantial similarity in actions within contests and mitigating concerns about heterogeneity in contest difficulty levels.[11,12] The symmetry property of the ν_{it} distribution can be used to implement the selection correction technique suggested by Pakes et al. (2015) and obtain upper bounds for the unobserved costs of

[11] Relative dispersion is defined as the ratio of the standard deviation to the mean number of submissions (Yoganarasimhan 2015).

[12] As a robustness check, I also tested for differences across contests offering different prize amounts by estimating the participant entry model separately for contests offering a total award of no more than $1000 and contests offering a total award greater than $1000. I did not find evidence of a difference in cost parameters (the identified sets overlapped).

non-entrants. As long as the number of entrants exceeds the number of non-entrants for a given contest, the negatives of the lowest lower bounds on cost unobservables over all participants can be used as upper bounds for the negatives of the cost unobservables of non-entrants.

For a given contest, the moment conditions can be developed as follows. First, rank all entrants by $gr_{it}^k = -g(z_{it}^k)\Delta r_{it}^*(d_{it}+1, d_{it}; \theta)$ so that $gr_{(1)t}^k \leq gr_{(2)t}^k \leq ... \leq gr_{(I_t)t}^k$. Next, construct a set of size equal to the number of non-entrants, such that $U_t^k = \{i : gr_{it}^k \geq gr_{(n_t+1)t}^k\}$, where n_t is the number of entrants in contest t. The negative lowest lower bounds for $\nu_{it} g(z_{it}^k)$ become the upper bounds for the $-\nu_{it} g(z_{it}^k)$ of non-entrants. Define the moment

$$m_k^U(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{\sqrt{I_t}}{I_t} \left(\sum_{i \in L_t} g(z_{it}^k)\Delta r_{it}(d_{it}, d_{it}-1; \theta) - \sum_{i \in U_t^k} g(z_{it}^k)\Delta r_{it}^*(d_{it}+1, d_{it}; \theta) \right),$$

where

$$\Delta r_{it}(d_{it}, d_{it}-1; \theta) = \Delta R_{it}(d_{it}, d_{it}-1) - \theta_1 - \theta_2(2d_{it}-1)$$

is the difference in observable profits from making one less submission.

Consider the expectational error $\omega_{itd_{it}}$. The lowest lower bounds on cost unobservables used as part of the selection correction technique originate from a selected subset of participants. I require an assumption on the joint density of expectational errors and cost unobservables to ensure that participants with the lowest costs do not consistently underestimate their expected marginal returns. Otherwise, the upper bounds I obtain for non-entrants may be too low. This assumption would only affect the observations used in constructing $m_k^U(\theta)$ for participants in U_t^k with $d_{it} < 5$ because the inequality condition for participants with $d_{it} = 5$ does not contain an expectational error term (Assumption 3). I find that this applies to less than 5% of all participant entry occasions and, as a result, does not have a consequential impact on estimated identified set of cost parameters. I provide the exact condition for the joint density of expectational errors and cost unobservables in Appendix A.3. The proof that if $m_k^U(\theta) \geq 0$, then θ lies in the identified set of cost parameters follows naturally from the proof presented in Pakes et al. (2015) and is reproduced in Appendix A.3 for completeness.

5.2.3 Identified Set

The identified set for parameters $\theta = (\theta_1, \theta_2)$ is defined as

$$\{\theta : m_k^L(\theta) \geq 0 \text{ and } m_k^U(\theta) \geq 0, \forall k\}.$$

Identification of the cost parameters follows naturally from the restrictions imposed by the moment inequalities. Note that a participant's action is a function of her cost unobservable, and hence, will necessarily be correlated with her cost unobservable, so that $E[\nu_{it}|d_{it}] \neq 0$. As a result, the instrument 1 can be used to partially identify the linear cost parameter θ_1, but additional instruments correlated with the participant's action are required to identify the quadratic cost parameter θ_2. I make the following assumption to obtain instruments for participant actions:

Assumption 5 *The cost unobservable ν_{it} is not correlated with* $\text{Past\_Earnings}_{it}$, Page_Views_{it} *and* Registration_i, *where* Registration_i *is a positive-valued function of participant i's registration date.*

I find it necessary to impose the stringent covariance restrictions on ν_{it} described in Assumption 5 to obtain an informative identified set. However, it is possible to relax these restrictions by making an assumption about the shape of the cost function instead.[13]

By Assumption 5, I can construct three instruments - 1, $\text{Past\_Earnings}_{it}$, and ($\text{Page\_Views}_{it} \times \text{Registration}_i$). This yields three lower bounds and three upper bounds for a total of six moment inequalities. To obtain the identified set, I search over a two-dimensional grid and accept all values that do not violate the six moment inequalities. It is possible that no θ satisfies all six moment inequalities, simply because of sampling error in the moments. I find this not to be the case and report an identified set for the θ_1 and θ_2 parameters.

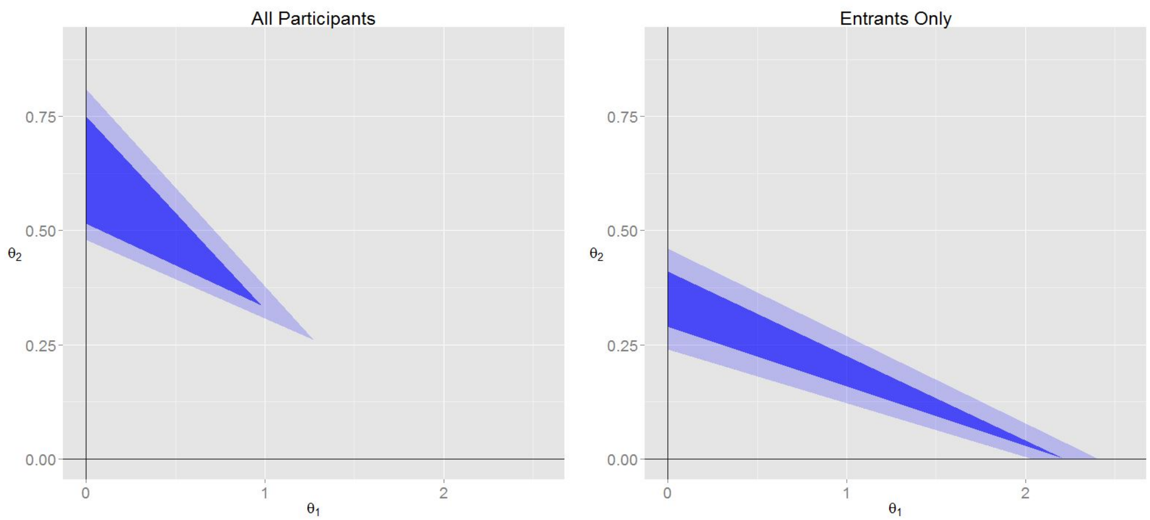
5.2.4 Cost Estimates

The left panel of Figure 4 presents a visualization of the identified set of cost parameters and Figure 5 shows the implied cost and marginal cost functions. The confidence set includes the true cost parameters 95% of the time and is obtained using a procedure suggested by Andrews and

[13] For example, Assumption 5 can be replaced with the assumption that $\theta_1 = 0$, so that $c_{it}(d_{it}) = \nu_{it} d_{it} + \theta_2 d_{it}^2$. In this case, I allow for flexibility in the covariance of cost unobservables and participant characteristics by restricting the shape of the cost function.

Soares (2010).[14] Intuitively, the procedure consists of simulating via bootstrap the distribution of a criterion function that penalizes violations of the moment inequalities. The simulated distribution is used to obtain a critical value, which is compared to the actual value of the criterion function in the observed sample. Points where the value of the criterion function falls below the critical value are included in the confidence set. The above procedure, first described by Chernozhukov et al. (2007), may produce very conservative confidence sets, primarily because of the influence of very positive moments that satisfy the inequality restrictions by a wide margin. Andrews and Soares (2010) suggest a moment selection procedure that yields more precise coverage by excluding very positive moments before simulating the criterion function.

Figure 4: Identified Sets for Cost Parameters



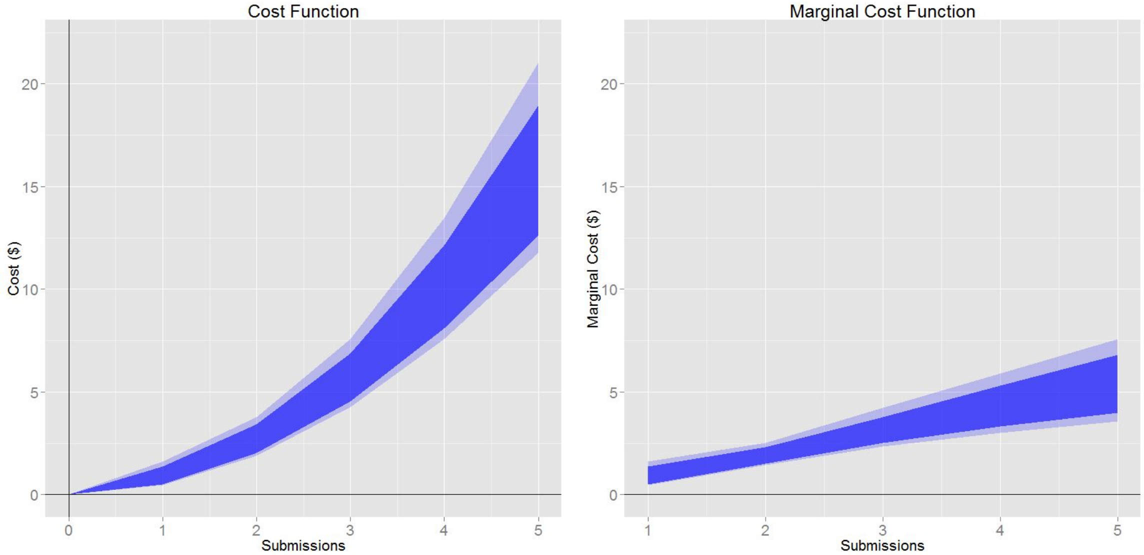
Note: Identified set $\{\theta : m_k^L(\theta) \geq 0$ and $m_k^U(\theta) \geq 0, \forall k\}$ is illustrated in the left panel. Right panel shows the identified set obtained using data only on entrants and ignoring selection. Parameters θ_1 and θ_2 restricted to be positive. 95% confidence sets obtained using Andrews and Soares (2010). First stage estimation error not taken into account. One contest excluded from estimation because number of non-entrants exceeded the number of entrants.

The estimated cost parameters suggest that on average, participants incur a cost of \$0.50-0.80 for producing the first submission. This cost estimate captures the cognitive and mental effort required to think of a 140 character idea as well as the opportunity cost of time that could have been spent elsewhere. Costs increase in a convex manner, with the average cost of making a fifth submission in the range of \$4-8. Participants require a substantial amount of additional effort to think of a large number of ideas.

For comparison, the right panel of Figure 4 shows the estimated identified set using data only

[14] The cost estimates do not take into account error from the first stage estimation of participant abilities. I found that estimating the full model for a re-sampled set of contests did not substantially change the cost estimates.

Figure 5: Identified Set of Cost Functions



Note: Identified set and confidence set of cost functions as implied by the cost parameters in the left panel of Figure 4 is illustrated.

on entrants and ignoring the issue of selection into contests. The upper bound of the identified set is noticeably lower than in the left panel, implying costs roughly half as large and possibly a smaller degree of convexity. This highlights the importance of taking selective entry into account. The estimates obtained using data on entrants only would overestimate the counterfactual response of entrants and ignore the possibility of submissions from non-entrants.

6 Counterfactuals

Although moment inequalities allow for flexible information sets in estimation, I require an explicit specification of the information sets of participants to simulate counterfactuals. I experiment with two specifications, which I refer to as *complete information* and *incomplete information*.

In the complete information setup, I assume that participants play a Nash equilibrium in submission strategies and know the prize structure of the contest, sponsor preferences, the number of competitors they face, their own characteristics, as well as competitor characteristics and actions. Formally, participant i's information set in contest t is given by $\mathcal{J}_{it}^{CI} = \{d_{it}, d_{-it}, X_{it}, X_{-it}, I_t, M_t, \beta_t\}$, where M_t represents the prize structure of contest t and includes the prize amount and number of prize spots. As before, I assume that participants know the density of quality shocks, F_ϵ. These assumptions enable the simulation of alternative designs at different values of parameters in the identified set and highlight the key incentives at work. For a uniformly sampled point in the identified set, I recover bounds on cost unobservables for each participant. These bounds ensure that at the sampled parameter vector, the observed decisions constitute an equilibrium. I uniformly sample cost unobservables that satisfy the bounds for each participant and compute equilibrium actions under alternative contest designs. I repeat the procedure for different sample parameters and cost draws, and recover bounds on the outcome of interest across simulations. Details of the counterfactual simulation procedure are provided in Appendix A.4.1.

The complete information assumption may require a high level of participant sophistication. Tongal reveals neither the identities nor the submissions of competitors. Furthermore, participants may not have a good sense of sponsor preferences. In the incomplete information scenario, I allow for participant uncertainty with regards to sponsor preferences and the quantity, characteristics and actions of competitors. Participant i's information set in contest t is given by $\mathcal{J}_{it}^{II} = \{d_{it}, X_{it}, M_t\}$, and the participant knows the density of quality shocks F_ϵ, the conditional joint density of the number of participants and sponsor preferences $H(I_t, \beta_t|M_t)$ and the conditional joint density of competitor actions and characteristics $G(d_{-it}, X_{-it}|I_t, M_t, \beta_t)$. I assume that participants play a Bayes-Nash equilibrium in submission strategies and that the same equilibrium is played in the data conditional on M_t. To simulate counterfactuals, I use an iterative procedure similar to Yoganarasimhan (2015), described further in Section 6.2 and Appendix A.4.2.

Equipped with participant ability estimates, cost functions, and an assumption about the information structure of the game, I run simulations of alternative contest designs. I focus on a number of outcome metrics. The total number of entrants $\sum_{i=1}^{I_t} 1\{d_{it} > 0\}$ and total submissions $\sum_{i=1}^{I_t} d_{it}$ are important metrics for the data provider. Increasing entry cultivates participant engagement with the platform and allows for sponsors to communicate with a large number of potential consumers and build brand awareness. Quality outcomes are also important if the goal of the sponsor is to implement the best idea or to incorporate information from all submitted ideas into its marketing strategy. I consider expected total quality (defined as ability-weighted submissions) $\sum_{i=1}^{I_t} e^{a_{it}} d_{it}$ and maximum quality $\log\left(\sum_{i=1}^{I_t} e^{a_{it}} d_{it}\right)$. A sponsor may be interested in total quality if it wishes to combine data from all submissions to create an ad or improve its product offerings. Maximum quality becomes more important for a sponsor interested in implementing only the best idea.

To develop intuition, I initially assume complete information and focus on a single contest organized by a large toy manufacturer. Section 6.3 presents the impact of the counterfactual policies across contests under both complete and incomplete information scenarios.

6.1 Designing a Contest

To demonstrate the effects of different design parameters, I focus on an ideation contest sponsored by a popular toy manufacturer and assume that participants have complete information. The contest offered 50 prizes of \$100 for a total award of \$5,000 and attracted a total of 254 participants, of whom 177 entered and made 565 submissions. Table 6 shows the impact of counterfactual policies on contest outcomes.

Table 6: Counterfactual Design Outcomes for a Sample Contest

	Entrants		Submissions		Total Quality		Max Quality	
	LB	UB	LB	UB	LB	UB	LB	UB
Actual	177		565		1		1	
Single Prize	169	182	551	564	1.00	1.01	1.00	1.00
20% Prize Increase	193	202	611	631	1.11	1.14	1.01	1.02
1 Submission Limit	238	253	238	253	0.35	0.35	0.87	0.87

Note: Expected total and maximum quality normalized to 1 for actual outcome. Participants assumed to have complete information.

6.1.1 Number of Prizes

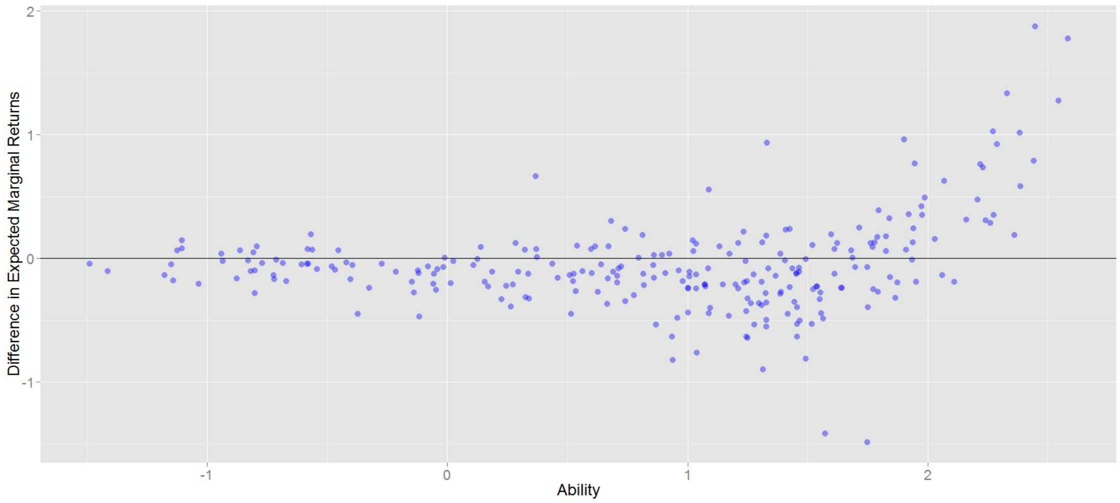
In this counterfactual, I reduce the total number of prizes to 1. Instead of offering 50 prizes of \$100, the sponsor now offers a single prize of \$5,000. The expected payoff of participant i can simply be written as

$$\pi_{it}^{\text{Single Prize}} = \frac{e^{a_{it}} d_{it}}{\sum_{j=1}^{I_t} e^{a_{jt}} d_{jt}} \times 5,000 - (\theta_1 + \nu_{it}) d_{it} - \theta_2 d_{it}^2. \tag{2}$$

Figure 6 illustrates the change in expected marginal returns experienced by participants of different ability levels in the contest. Consistent with the structural model, higher ability participants prefer a single prize and experience an increase in expected marginal returns, whereas lower ability participants prefer multiple prizes and experience a decrease in expected marginal returns. Higher ability participants may submit more, whereas lower ability participants may submit less.

The row of Table 6 labeled "Single Prize" illustrates the impact of reducing the number of prizes on contest outcomes. I compare the simulated outcomes to the outcomes observed in the data. Overall, the effects are not very large. This can be explained as follows. An increase in

Figure 6: Impact of Reducing Number of Prizes on Expected Marginal Returns

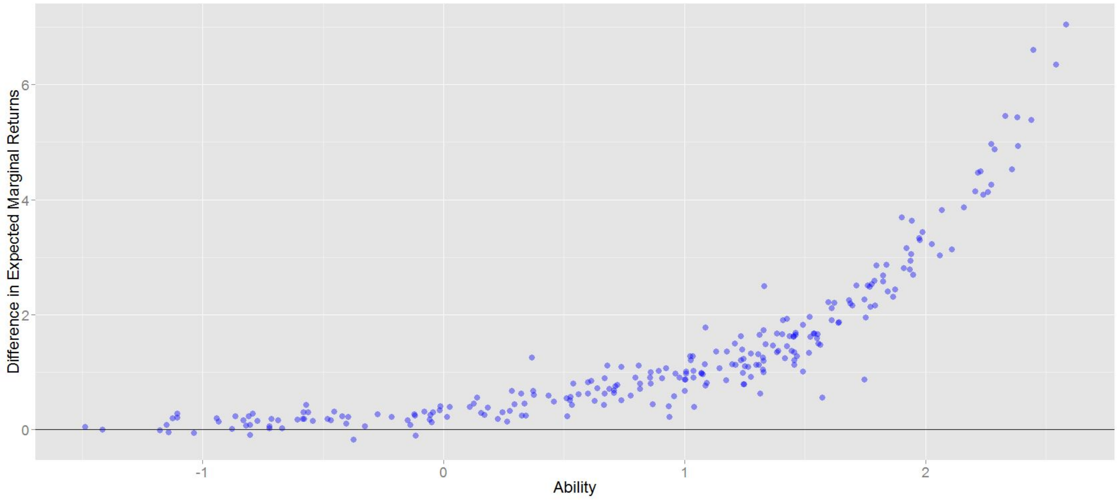


submissions from higher ability participants is balanced out by a decrease in submissions from lower ability participants. Also, few participants alter their actions because the expected marginal returns from doing so are low (on the order of $2 for the highest ability entrant). As higher ability participants already tend to make multiple submissions (recall Figure 3), the cost of making an additional submission likely exceeds the expected return. This result suggest that for contests that attract a substantial number of submissions, the number of prizes may not significantly affect the contest outcome, holding fixed total award.

6.1.2 Prize Amount

Consider the impact of increasing the total prize award. I use Equation 2 to simulate a 20% increase in total award and report the impact relative to the outcome of a single prize contest. Figure 7 displays the impact of the policy on the expected marginal returns of participants. As expected, all participants appear to prefer a larger prize, although higher ability participants reap more benefits from it. The row of Table 6 labeled "20% Prize Increase" illustrates the counterfactual impact on outcome metrics. I observe a significant increase in the number of entrants (↑9-14%), submissions (↑8-12%) and expected total quality (↑11-14%). Higher ability participants predominantly account for the increase in expected total quality. Expected maximum quality does not change substantially (↑1-2%), which is not surprising given that the contest already attracts a large number of submissions.

Figure 7: Impact of 20% Prize Increase on Expected Marginal Returns



6.1.3 Entry Limits

The platform requires that all participants can submit at most five times to each contest. What if participants could submit at most once? Higher ability participants would be restricted by a lower submission limit as they tend to make more submissions (recall Figure 3). The final row of Table 6 shows that the submission limit increases the number of entrants quite significantly (↑34-43%) but naturally reduces the number of submissions (↓55-58%). In particular, every participant who considered the contest chooses to enter but submits only one idea. Higher ability participants with low costs no longer crowd out other potential entrants. However, by significantly reducing the number of submissions, the platform reduces expected total quality (↓65%) and expected maximum quality (↓13%). Although high ability entrants who submitted more than once can no longer submit as much, high ability non-entrants with high costs can now enter the contest, together with low ability non-entrants, which may potentially balance out the impact on expected quality outcomes. However, as the set of non-entrants for this contest consisted mostly of lower ability participants, a more stringent submission limit resulted in higher entry but a larger proportion of lower quality submissions, reducing expected total and maximum quality.

6.2 The Impact of Incomplete Information

To simulate counterfactuals under incomplete information, it is necessary to recover $H(I_t, \beta_t | M_t)$ and $G(d_{-it}, X_{-it} | I_t, M_t, \beta_t)$, which can theoretically be achieved by flexible density estimation. However, I find this to be infeasible given the large number of contest-specific variables. Instead, I focus on a subset of 49 contests that offered four \$250 prizes and treat each contest as an independent draw from the joint density of sponsor preferences, the number of competitors, and competitor actions and characteristics conditional on the contest structure. All incomplete information counterfactual analyses are conducted only for this subset of contests, labeled $\mathcal{W}$.

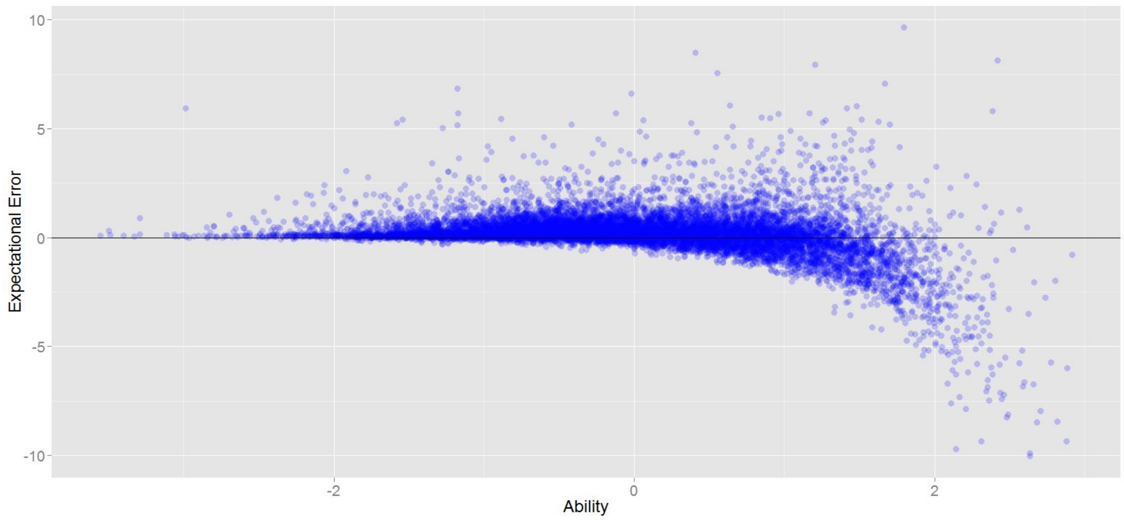
To understand the impact of incomplete information on behavior, I recover participant expectational errors, which capture the difference between a participant's expected returns under incomplete information and her expected returns under complete information. To do so, I draw a sample of contests of size B from $\mathcal{W}$ (with replacement) and label these contests $b = 1, ..., B$. I let $\{d_{-ib}, X_{-ib}, I_b, \beta_b\}$ denote the variables associated with contest b and let $a_{it}^b = \beta_b X_{it}$. Then, assuming that the cost unobservables ν_{it} are independent conditional on X_{it} for all participants i within a contest t and letting j_b denote a random participant in contest b, the expected returns $E\left[R_t(d_{it}, d_{-it}; a_{it}, a_{-it}) | \mathcal{J}_{it}\right]$ can be approximated by

$$ER_{it}(d_{it}) = \frac{1}{B} \sum_{b=1}^{B} R_t(d_{it}, d_{-j_b b}; a_{it}^b, a_{-j_b b})$$

for $t \in \mathcal{W}$. This is akin to assuming that the participant knows the variables associated with each contest in $\mathcal{W}$, but does not know which one of these contests she is playing. An estimate of participant $i's$ expectational error is given by $\hat{\omega}_{itd_{it}} = ER_{it}(d_{it}) - R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$.

Figure 8 illustrates the relationship between expectational errors and true participant abilities. Higher ability participants tend to underestimate their expected returns as they do not know with certainty that they are the highest ability participants in their contests. Similarly, lower ability participants tend to overestimate their expected returns as they do not know that they fall in the lower range of abilities within the contests they participate in. Given their observed actions, this implies that higher ability participants will have lower estimated costs than had they had complete information. Similarly, lower ability participants will have higher estimated costs than in a complete information scenario. As a result, the impact of incomplete information on counterfactual outcomes is not evident. A simulation analysis must be conducted to study the effects of this alternative assumption.

Figure 8: Joint Distribution of Abilities and Expectational Errors



Note: Abilities standardized within contests. Standard deviation of expectational errors normalized to one within contests.

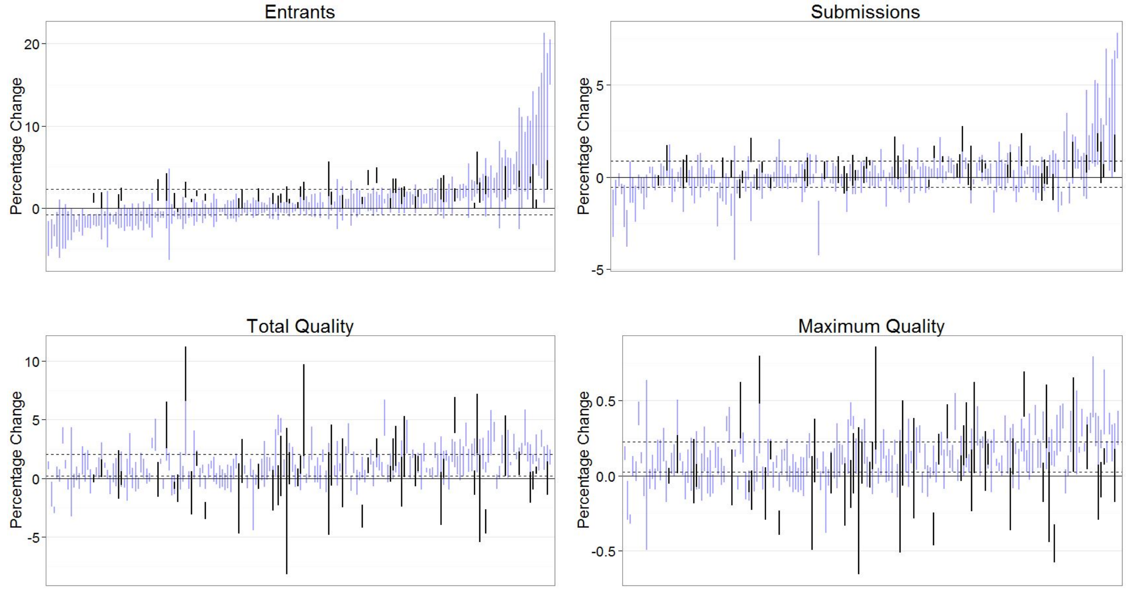
6.3 Counterfactual Outcomes Across Contests

I obtain bounds on the outcomes of all contests for each one of the three design counterfactuals under the assumption of complete information. For the 49 contests in $\mathcal{W}$, I also obtain counterfactual outcomes under the assumption of incomplete information using an iterative procedure described in Appendix A.4.2.[15] The intuition presented in Section 6.1 carries over naturally and can be used to explain the directional impact of the design parameters. A sponsor interested in organizing an ideation contest can draw on these results to understand the expected implications of different design decisions. Figures 9, 10 and 11 show the impact of counterfactual design policies on different contests, and Table 7 shows the average impact across contests.

For most of the contests, reducing the number of prizes while holding fixed total award does not have a substantial impact on outcomes because the change in expected marginal returns to participants is low, meaning that few participants will alter their actions. Furthermore, of the

[15]This procedure finds one of possibly many counterfactual equilibria for a given parameter vector. However, in complete information simulations, I find that if multiple equilibria do exist, participant actions do not differ significantly across equilibria.

Figure 9: Impact of Reducing the Number of Prizes Across Contests

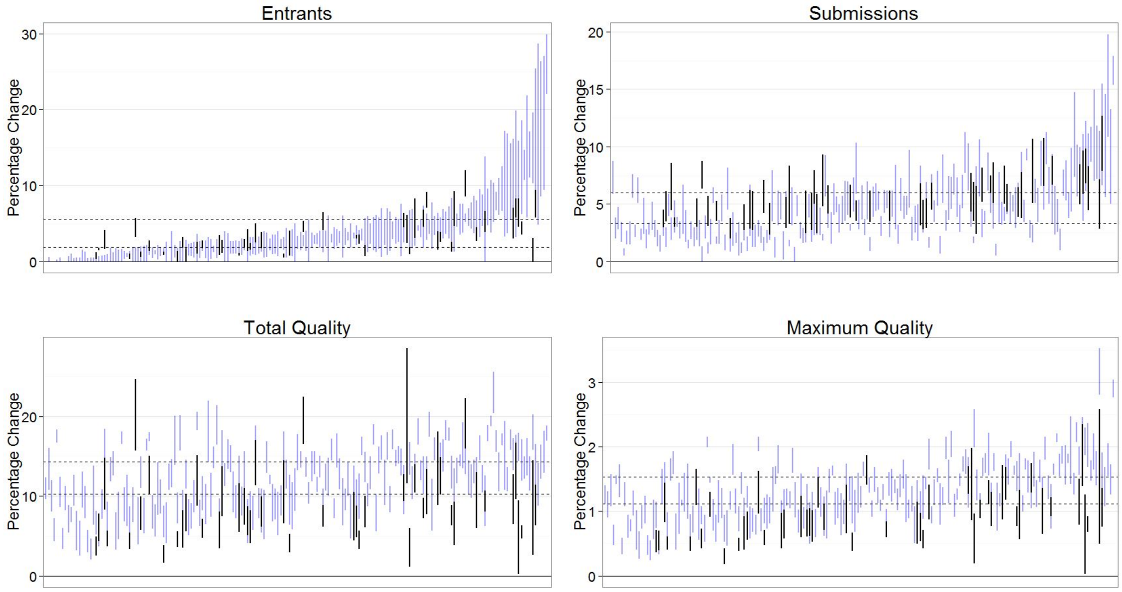


Note: Each segment represents the range of counterfactual outcomes for a single contest under complete information (light) and incomplete information (dark). Contests ordered by increasing impact on the number of entrants in the complete information scenario.

participants who do change their actions, higher ability participants tend to submit more, but lower ability participants tend to submit less, balancing out the overall impact on outcomes. For certain contests with very low dispersion in participant abilities, a single prize may significantly increase entry, but only under the assumption of complete information. In a setting with limited ability heterogeneity, there is no longer a reason to motivate lower ability participants, and a single prize reduces the incentive for all participants to rank lower. However, if participants do not know that they are in a contest with low ability heterogeneity, they do not react as strongly to a reduction in the number of prizes. A 20% prize increase improves the outcome metrics, especially expected total quality, but may not lead to as significant an increase in entry in a complete information scenario if there is substantial ability dispersion. The added incentive encourages higher ability participants to submit more, limiting the increase in expected returns from additional submissions for the remaining participants. A more stringent submission limit encourages entry but restricts higher ability participants, reducing expected total and maximum quality across both complete and incomplete information scenarios.

Overall, the average outcomes under complete information correspond very closely to the average outcomes under incomplete information. However, contest specific outcomes may differ substan-

Figure 10: Impact of a 20% Prize Increase Across Contests



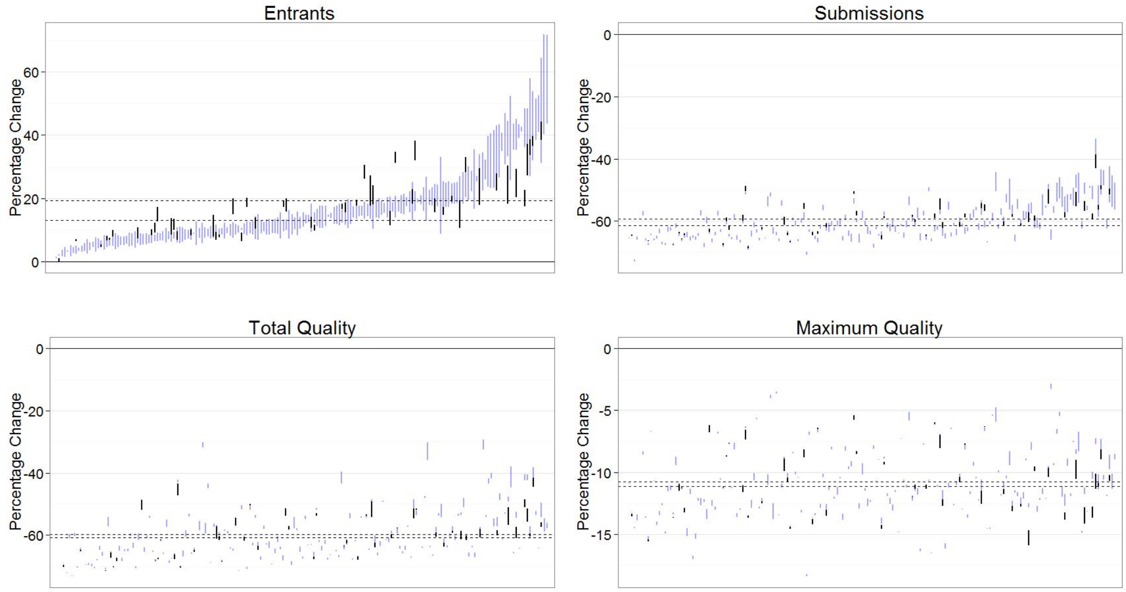
Note: Each segment represents the range of counterfactual outcomes for a single contest under complete information (light) and incomplete information (dark). Contests ordered by increasing impact on the number of entrants in the complete information scenario.

tially across the two informational assumptions. The impact of a design parameter on a contest outcome depends critically on the dispersion of participant abilities and costs within the contest. With complete information, participants accurately recover the distributions of abilities and costs within a contest and act accordingly. With incomplete information, participants do not know sponsor preferences and the characteristics of their competitors. As a result, they cannot correctly infer the extent of ability and cost dispersion within a contest and take different actions than they would have in a complete information scenario.

6.3.1 Validation

To check whether the structural model yields reasonable predictions, I compare the counterfactual simulation outcomes to the predictions of the fixed effects regressions from Section 3.1 and Appendix A.2. The regression coefficients suggest that a 20% prize increase leads to a 5.5-11.8% increase in participant submissions across the contests in $\mathcal{W}$. Simulations based on the structural model suggest that participants increase submissions by 2.9-5.7% under complete information and by 4.0-7.1% under incomplete information. Despite different underlying assumptions, both the structural model and the fixed effects regressions yield similar predictions of how submissions respond to

Figure 11: Impact of a 1 Submission Limit Across Contests



Note: Each segment represents the range of counterfactual outcomes for a single contest under complete information (light) and incomplete information (dark). Contests ordered by increasing impact on the number of entrants in the complete information scenario.

changes in prize amount, lending credibility to the counterfactual outcomes obtained through the structural model.

6.3.2 Practical Implications

These findings have a number of practical implications for ideation contest design. First, the choice of how many prizes to offer does not substantially affect the outcome of the contest, as long as the contest attracts a large number of submissions. Furthermore, I do not expect alternative prize distributions to significantly affect outcomes. The difference in outcomes between an even prize distribution and a prize distribution that allocates 100% of the prize money to the first prize is small. As a result, any intermediate prize allocation must have less of an effect on participant expected marginal returns, and hence, less of an impact on contest outcomes. The choice of how many prizes to offer should be driven by institutional considerations. For example, in many ideation contests, the sponsor retains intellectual property of the winning submissions. In these settings, the sponsor would benefit from offering more prizes, without significantly altering the outcome of a contest. Only in complete information settings where the sponsor expects to receive submissions from a very homogeneous set of participants does a single prize appear preferable. Second, if a

Table 7: Average Counterfactual Design Outcomes Across Contests

	Entrants		Submissions		Total Quality		Max Quality	
	LB	UB	LB	UB	LB	UB	LB	UB
Complete Information (all contests)								
Single Prize	-0.8	2.2	-0.5	0.9	0.2	2.0	0	0.2
20% Prize Increase	1.9	5.5	3.3	6.0	10.3	14.2	1.1	1.5
1 Submission Limit	13.0	19.2	-61.3	-59.2	-60.8	-59.6	-11.1	-10.7
Incomplete Information (49 contests offering four $250 prizes)								
Single Prize	0.9	2.7	-0.1	1.0	-1.0	2.4	-0.1	0.2
20% Prize Increase	2.4	4.3	4.0	7.1	6.5	11.5	0.7	1.2
1 Submission Limit	15.4	19.5	-60.1	-58.6	-60.8	-59.2	-11.1	-10.6

Note: Average lower bound (LB) and upper bound (UB) of percentage change in counterfactual outcomes reported.

sponsor's intention is to increase the number of entrants, increasing prize award may not have as strong an impact if there is no limit on the maximum number of submissions a participant can make. Third, a submission limit can be used as an effective strategy to encourage entry but may come at the cost of expected total and maximum idea quality. If a sponsor seeks to attract a large number of entrants and use the contest as a mechanism for engaging potential consumers, it should implement a more stringent submission limit.

7 Conclusion

Firms across a range of industries use ideation contests to procure ideas for ads, new products and marketing strategies. An appropriate design can improve the outcome of a contest. Moreover, different firms may care about different outcome metrics. Brands interested in engaging consumers may focus on increasing entry, whereas a manufacturer interested in designing a new product may value the maximum quality of submitted ideas.

I empirically investigate the impact of three design parameters - number of prizes, prize amount and submission limit - on contest participation and quality outcomes, using data from a popular crowdsourcing studio that runs ad ideation contests for major brands. I present a structural model of an ideation contest that allows for heterogeneity in participant abilities and costs. Participants may select into contests based on an unobservable component of costs and may have incomplete information. Participant abilities are estimated off of sponsor ranking data using a hierarchical Bayesian methodology, and costs are partially identified from participant submission decisions using moment inequalities. Counterfactual simulations reveal the impact of different designs. The results show that the number of prizes does not significantly affect contest outcomes, prize amount increases submissions and all expected quality metrics but may not necessarily increase entry, and submission limits encourage entry but significantly reduce expected total and maximum quality.

I make several simplifying assumptions to ensure the model remains feasible. First, I assume that each contest is an independent game. Participants face no dynamic incentives and do not have constraints that prevent them from entering multiple contests at the same time. Future research may examine the implications of dynamics and competing contests on the optimal design of contest platforms. Second, as my sample is limited to 140 character ad ideation contests, I assume that participants cannot choose to alter the quality of an individual submission but are instead endowed with an ability that determines the average quality of their submissions. Future research may examine the optimal design of ideation contests for more complicated tasks and allow for participants to choose not only how many submissions to make but also how much effort to invest into each individual submission. Finally, I do not observe the actual applications of ideas obtained through the contests. Absent these data, quality is inferred from sponsor rankings of winners and represents the sponsor's but not necessarily the market's perception of idea quality. Future work may incorporate post-contest outcomes to assess the impact of contest design on the "applicability" of winning ideas.

An appropriately designed ideation contest can yield interesting ideas and spur innovation. Crowdsourcing platforms, contest sponsors and designers need to carefully consider the effects of different design parameters on contest outcomes.

References

Andrews, Donald WK, Gustavo Soares. 2010. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78**(1) 119–157.

Baye, Michael R, Dan Kovenock, Casper G De Vries. 1993. Rigging the lobbying process: an application of the all-pay auction. *American Economic Review* 289–294.

Baye, Michael R, Dan Kovenock, Casper G De Vries. 1994. The solution to the tullock rent-seeking game when $r > 2$: Mixed-strategy equilibria and mean dissipation rates. *Public Choice* **81**(3-4) 363–380.

Beggs, Steven, Scott Cardell, Jerry Hausman. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics* **17**(1) 1–19.

Borkovsky, Ron N, Paul B Ellickson, Brett R Gordon, Victor Aguirregabiria, Pedro Gardete, Paul Grieco, Todd Gureckis, Teck-Hua Ho, Laurent Mathevet, Andrew Sweeting. 2015. Multiplicity of equilibria and information structures in empirical games: challenges and prospects. *Marketing Letters* **26**(2) 115–125.

Boudreau, Kevin J, Nicola Lacetera, Karim R Lakhani. 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science* **57**(5) 843–863.

Boudreau, Kevin J, Karim Lakhani, Michael E Menietti. 2015. Performance responses to competition across skill-levels in rank order tournaments: Field evidence and implications for tournament design. *RAND Journal of Economics* .

Chapman, Randall G, Richard Staelin. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of marketing research* 288–301.

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

Che, Yeon-Koo, Ian L Gale. 1998. Caps on political lobbying. *American Economic Review* 643–651.

Chernozhukov, Victor, Han Hong, Elie Tamer. 2007. Estimation and confidence regions for parameter sets in econometric models1. *Econometrica* **75**(5) 1243–1284.

Dahan, Ely, Haim Mendelson. 2001. An extreme-value model of concept testing. *Management Science* **47**(1) 102–116.

Dohmen, Thomas, Armin Falk. 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review* 556–590.

Eriksson, Tor, Sabrina Teyssier, Marie-Claire Villeval. 2009. Self-selection and the efficiency of tournaments. *Economic Inquiry* **47**(3) 530–548.

Fullerton, Richard L, R Preston McAfee. 1999. Auctioning entry into tournaments. *Journal of Political Economy* **107**(3) 573–605.

Girotra, Karan, Christian Terwiesch, Karl T Ulrich. 2010. Idea generation and the quality of the best idea. *Management Science* **56**(4) 591–605.

Gross, Daniel P. 2014. Creativity under fire: The effects of competition on innovation and the creative process. *Working Paper* .

Ishii, Joy. 2008. Compatibility, competition, and investment in network industries: ATM networks in the banking industry. *Working Paper* .

Kalra, Ajay, Mengze Shi. 2001. Designing optimal sales contests: A theoretical perspective. *Marketing Science* **20**(2) 170–193.

Kirby, Julia. 2013. Creative that cracks the code. *Harvard Business Review* **91**(3) 86–89.

Klaassen, Abbey. 2011. Harley-Davidson breaks consumer-created work from Victors & Spoils. URL `http://adage.com/article/news/harley-davidson-breaks-consumer-created-victors-spoils-ad/148873/`. [Online; posted 14-February-2011].

Lazear, Edward P, Sherwin Rosen. 1981. Rank-order tournaments as optimum labor contracts. *The Journal of Political Economy* **89**(5) 841–864.

Loch, Christoph H, Christian Terwiesch, Stefan Thomke. 2001. Parallel and sequential testing of design alternatives. *Management Science* **47**(5) 663–678.

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

Pakes, A., J. Porter, Kate Ho, Joy Ishii. 2015. Moment inequalities and their application. *Econometrica* **83**(1) 315–334.

Pickering, Ben. 2011. 4 reasons why contests should be part of your marketing strategy. URL http://mashable.com/2011/07/21/contest-marketing/. [Online; posted 21-July-2011].

Rossi, Peter E, Greg M Allenby, Rob McCulloch. 2012. *Bayesian statistics and marketing*. John Wiley & Sons.

Sisak, Dana. 2009. Multiple-prize contests–the optimal allocation of prizes. *Journal of Economic Surveys* **23**(1) 82–114.

Stein, William E. 2002. Asymmetric rent-seeking with more than two contestants. *Public Choice* **113**(3-4) 325–336.

Szymanski, Stefan, Tommaso M Valletti. 2005. Incentive effects of second prizes. *European Journal of Political Economy* **21**(2) 467–481.

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

Terwiesch, Christian, Yi Xu. 2008. Innovation contests, open innovation, and multiagent problem solving. *Management Science* **54**(9) 1529–1543.

Toubia, Olivier. 2006. Idea generation, creativity, and incentives. *Marketing Science* **25**(5) 411–425.

Tullock, Gordon. 1980. Efficient rent seeking. *Efficient Rent-Seeking*. Springer, 3–16.

Wooten, Joel O, Karl T Ulrich. 2011. Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments. *Working Paper* .

Wooten, Joel O, Karl T Ulrich. 2013. The impact of visibility in innovation tournaments: Evidence from field experiments. *Working Paper* .

Yoganarasimhan, Hema. 2015. Estimation of beauty contest auctions. *Marketing Science* .

A Appendix

A.1 Summary Statistics for Participant Characteristics

I use data on the past behavior and self-reported demographics of participants to construct a set of participant characteristics. Table 4 in the main text defines the set of variables used throughout the analysis. In this section, I present summary statistics.

Table 8 presents the minimum, median, mean and maximum values for each continuous characteristic and Table 9 summarizes the variation in the discrete characteristics. In all cases, an observation is a participant-contest pair, of which there are a total of 44,743 (35,011 cases of entry and 9,732 cases of non-entry).

Table 8: Summary Statistics for Continuous Participant Characteristics

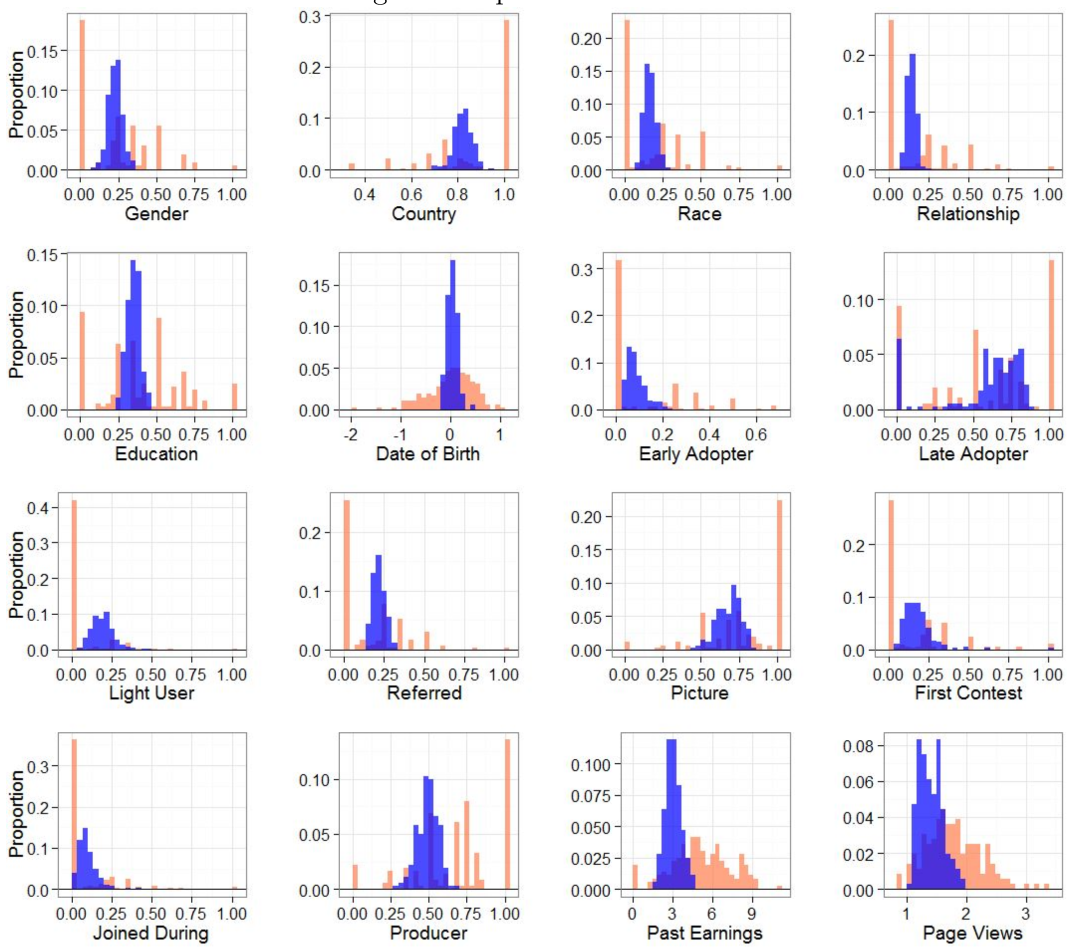
Variable	Minimum	Median	Mean	Maximum
Demographics				
Date of Birth	-3.154	0.268	0.045	1.986
	1950	*1984*	*1982*	*2001*
Skill and Interest				
Past Earnings	0	0	2.773	12.816
	$0	*$0*	*$15*	*$367,925*
Page Views	0.693	1.386	1.411	6.909
	1	*3*	*3*	*1,000*

Table 9: Summary Statistics for Discrete Participant Characteristics

Variable	0	1
Demographics		
Gender	34,335	10,408
Country	7,895	36,848
Race	37,385	7,358
Relationship	38,522	6,221
Education	29,359	15,384
Participant-Platform Characteristics		
Early Adopter	41,138	3,605
Late Adopter	17,193	27,550
Light User	35,763	8,980
Referred	35,118	9,625
Picture	14,730	30,013
Skill and Interest		
First Contest	36,307	8,436
Joined During	40,223	4,520
Producer	22,947	21,796

For estimation of the sponsor choice model, I use monotonic transformations of the continuous variables (described in Table 4). The date of birth variable is standardized to make the scale of the associated coefficient similar to the scale of the other coefficients. I apply log-transformations to past earnings and page views to mitigate the impact of very high-valued observations and account for the possibility of decreasing returns to skill or interest in the contest. I report summary statistics for the untransformed variables in italics. The past earnings summary statistics are rescaled by a constant to preserve participant anonymity. Overall, there is substantial variation in participant characteristics as less than 2% of all participants share the same set of characteristics.

Figure 12: Distributions of Average Participant and Winner Characteristics Across Contests



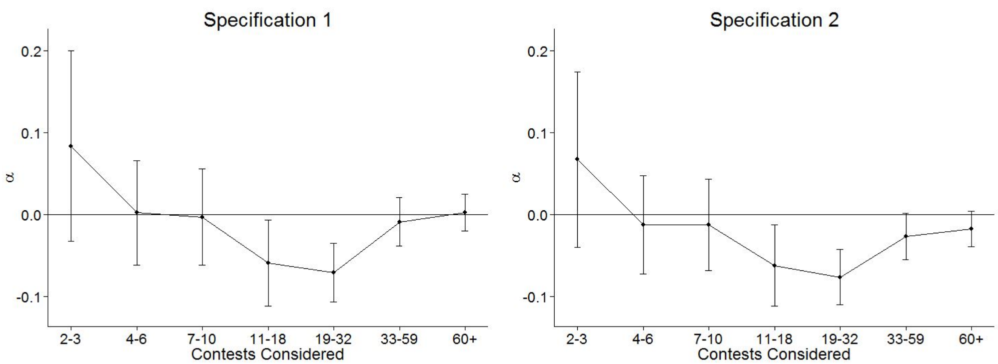
Note: Plot shows distribution of contest-specific averages for each characteristic in X_{it} for participants who did not win (dark) and for participants who won (light).

For each characteristic, Figure 12 shows the distribution of the characteristic's average value across contests for all participants who did not win (dark) and for all participants who won (light). The plots imply that many contests attract a very similar set of participants. However, winner characteristics may differ substantially across contests, suggesting that if a certain type of participant appears to have an advantage in one contest, she may not necessarily have an advantage in a different contest. For example, contests exhibit significant dispersion in the average winner's education and gender, whereas all contests appear to attract a similar proportion of genders and education levels.

A.2 Further Descriptive Evidence

I illustrate that the regression findings in Section 3.1 are robust to different specifications of the regression equation and different classifications of heavy and light users. I classify participants into groups based on the number of contests they considered.[16] Each group contains a similar number of participants-contest decision instances. Figure 13 illustrates the value of the coefficient α when Regression 1 is applied separately to each group of participants, where X_t is defined as the number of prizes offered in contest t. The regression in the left panel does not allow for contest-varying participant characteristics, whereas the regression in the right panel does.

Figure 13: Effects of Number of Prizes on Submissions



Note: Plots show estimates and 95% confidence intervals for the effect of the number of prize spots on submissions. Participants grouped based on the number of contests they entered. Specification 1 is $Y_{it} = \alpha X_t + \xi_{iC_t} + \epsilon_{it}$*. Specification 2 is* $Y_{it} = \alpha X_t + \gamma Z_{it} + \xi_{iC_t} + \epsilon_{it}$*.*

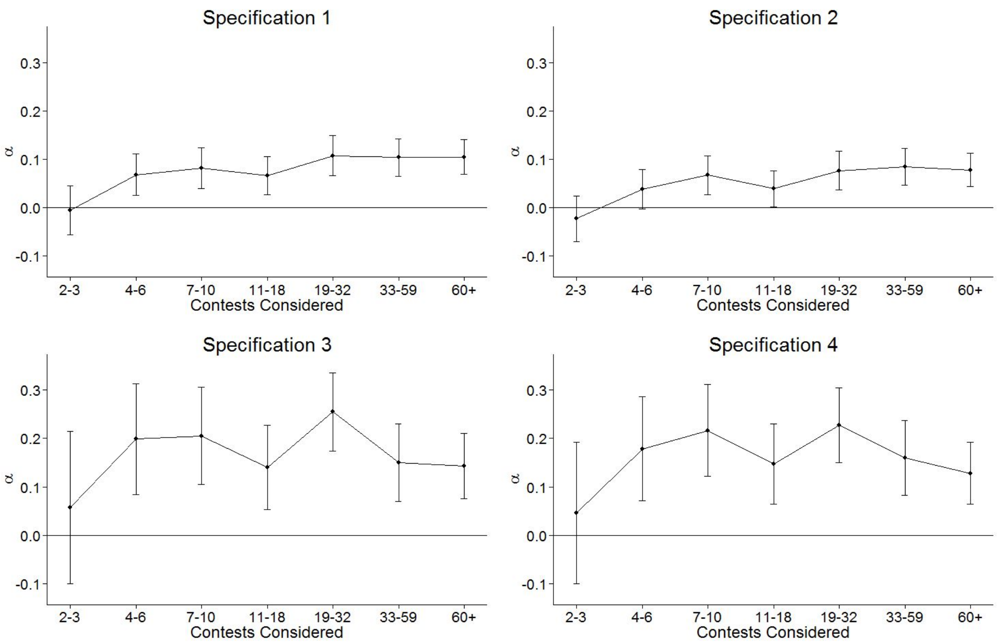
Participants who considered entering between 11 and 32 contest show evidence of a distaste for multiple prizes. However, this effect disappears for the most frequent participants. Further inspection reveals that participants who consider over 33 contests tend to submit near the maximum number of times to each contest, perhaps because of low costs or high abilities. As a result, there is limited variation in their submission behavior, resulting in a near-zero coefficient for the most frequent participants. Overall, Figure 13 shows evidence consistent with the theoretical prediction that lower cost or higher ability participants may prefer fewer prize, holding fixed total award.

Figure 14 similarly illustrates the estimates of α for the modified version of Regression 1 that

[16]The number of contests a participant considers is highly correlated with the number of contests she enters.

specifies X_t as the total award in contest t. The four panels illustrate the coefficient estimates obtained under different fixed effects structures and covariates. Consistent with theoretical predications, the plots show evidence that participants who enter more contests, perhaps because of lower costs or higher abilities, tend to react more to changes in prize amount.

Figure 14: Effects of Prize Amount on Submissions



Note: Plots show estimates and 95% confidence intervals for the effect of the prize amount on submissions. Participants grouped based on the number of contests they entered. Specification 1 is $Y_{it} = \alpha X_t + \xi_i + \epsilon_{it}$. *Specification 2 is* $Y_{it} = \alpha X_t + \gamma Z_{it} + \xi_i + \epsilon_{it}$. *Specification 3 is* $Y_{it} = \alpha X_t + \xi_{iC_t} + \epsilon_{it}$. *Specification 4 is* $Y_{it} = \alpha X_t + \gamma Z_{it} + \xi_{iC_t} + \epsilon_{it}$.

A.3 Deriving an Upper Bound on Marginal Costs

In this section, I reproduce the proof presented in Pakes et al. (2015), adapted to my notation and setting, to show that $m_k^U(\theta) \geq 0$. For clarity, I drop the t subscript and focus on a single contest, and use $g(z_{it}^k) = 1$. The proof can be amended to incorporate other instruments.

First, let $\Delta r_i = -\Delta r_i^*(d_i + 1, d_i; \theta)$ and use order-statistic notation to rank participants by ν_i and Δr_i, so that $\nu_{(1)} \leq \nu_{(2)} \leq ... \leq \nu_{(I)}$ and $\Delta r_{(1)} \leq \Delta r_{(2)} \leq ... \leq \Delta r_{(I)}$. Then, define the sets $L = \{i : d_i > 0\}$, $L_\nu = \{i : \nu_i \leq \nu_{(n)}\}$, $U = \{i : \Delta r_i \geq \Delta r_{(n+1)}\}$, and $U_\nu = \{i : \nu_i \leq \nu_{(I-n)}\}$, where I is the total number of participants, and n is the number of entrants. Let the change in expected profits from making $d_i - 1$ to d_i submissions for $i \in L$ be

$$\Delta\pi_i(d_i, d_i - 1) = \Delta r_i(d_i, d_i - 1; \theta) - \nu_i + \omega_{id_i,d_i-1}$$

and similarly, let the change in expected profits from making one additional submission be

$$\Delta\pi_i(d_i + 1, d_i) = \Delta r_i^*(d_i + 1, d_i; \theta) - \nu_i + \omega^*_{id_i+1,d_i}$$

where $\omega^*_{id_i+1,d_i} = \omega_{id_i+1,d_i}$ if $d_i < 5$ and $\omega^*_{id_i+1,d_i} = 0$ otherwise. Then, we have that

$$\begin{aligned}
&\frac{1}{I}\sum_{i\in L}\Delta r_i(d_i, d_i - 1; \theta) - \frac{1}{I}\sum_{i\in U}\Delta r_i^*(d_i + 1, d_i; \theta) \\
&\qquad \geq \frac{1}{I}\sum_{i\in L}\Delta r_i(d_i, d_i - 1; \theta) - \frac{1}{I}\sum_{i\in U_\nu}\Delta r_i^*(d_i + 1, d_i; \theta) \\
&\qquad = \frac{1}{I}\sum_{i\in L}\left(E[\Delta\pi_i(d_i, d_i - 1)|\mathcal{J}_i] + \nu_i - \omega_{id_i,d_i-1}\right) \\
&\qquad - \frac{1}{I}\sum_{i\in U_\nu}\left(E[\Delta\pi_i(d_i + 1, d_i)|\mathcal{J}_i] + \nu_i - \omega^*_{id_i+1,d_i}\right) \\
&\qquad\qquad \geq \frac{1}{I}\left(\sum_{i\in L}\nu_i - \sum_{i\in U_\nu}\nu_i\right) - \frac{1}{I}\left(\sum_{i\in L}\omega_{id_i,d_i-1} - \sum_{i\in U_\nu}\omega^*_{id_i+1,d_i}\right),
\end{aligned}$$

where the first inequality follows from the definition of the set U. The second inequality follows from the assumption that participants take the optimal action given their information sets. Note that

$$\frac{1}{I}\left(\sum_{i\in L}\nu_i - \sum_{i\in U_\nu}\nu_i\right) \geq \frac{1}{I}\left(\sum_{i\in L_\nu}\nu_i - \sum_{i\in U_\nu}\nu_i\right) = \frac{1}{I}\left(\sum_{i=1}^{n}\nu_{(i)} - \sum_{i=1}^{I-n}\nu_{(i)}\right).$$

The distributional assumption on ν_i (Assumption 4) ensures that

$$E\left[\frac{1}{I}\left(\sum_{i=1}^{n}\nu_{(i)} - \sum_{i=1}^{I-n}\nu_{(i)}\right)\right] \geq 0.$$

Furthermore,

$$E\left[\frac{1}{I}\sum_{i\in L}\omega_{id_i,d_i-1}\right] = \frac{1}{I}\sum_{i=1}^{I}E\left[1\{d_i > 0\}\omega_{id_i,d_i-1}\right] = \frac{1}{I}\sum_{i=1}^{I}E\left[1\{d_i > 0\}E[\omega_{id_i,d_i-1}|\mathcal{J}_i]\right] = 0.$$

Expectational errors are mean-zero for entrants because the action d_i is an element of the participant's information set. I also require the following assumption:

Assumption 6 $E\left[\frac{1}{I}\sum_{i\in U_\nu}\omega^*_{id_i+1,d_i}\right]\geq 0.$

In other words, participants in U_ν cannot consistently underestimate their expected marginal returns. Note that this applies only to participants in U_ν with $d_i<5$, as otherwise, $\omega^*_{id_i+1,d_i}=0$.

As a result,

$$E\left[\frac{1}{I}\sum_{i\in L}\Delta r_i(d_i,d_i-1;\theta)-\frac{1}{I}\sum_{i\in U}\Delta r^*_i(d_i+1,d_i;\theta)\right]\geq 0.$$

The empirical analogue $m^U_k(\theta)$ may incorporate instruments. A slight modification to the proof presented above can be used to show that if $m^U_k(\theta)\geq 0$, then θ is an element of the identified set. See the appendix of Pakes et al. (2015) for a proof that incorporates instruments.

A.4 Counterfactual Simulation Procedure

A.4.1 Complete Information

To simulate counterfactual contest designs under complete information, I draw sample parameters from the identified set and use iterated best response to obtain equilibrium strategies. I make the assumption that first-stage ability estimates are obtained without error. The following steps can be used to obtain counterfactual equilibrium outcomes for a contest t:

1. Uniformly sample θ^s from the identified set of average cost parameters.

2. At the sampled parameter, obtain bounds on the cost draw for each participant. Note that if θ^s were the true parameter, then by revealed preference, $\nu_{it}\geq\Delta r^*_{it}(d_{it}+1,d_{it};\theta^s)$ at the observed submission decisions, where $\Delta r^*_{it}(d_{it}+1,d_{it};\theta^s)$ is evaluated at the sampled parameter θ^s and estimated abilities a_{it} and a_{-it}. Similarly, $\nu_{it}\leq\Delta r_{it}(d_{it},d_{it}-1;\theta^s)$ if participant i submitted at least once to contest t. Otherwise, I use $\nu_{it}\leq\max_{j=1,\dots,I_t}\{-\Delta r^*_{jt}(d_{jt}+1,d_{jt};\theta^s)\}$ as an approximation to the upper bound. For each participant, obtain a lower bound ν^{Ls}_{it} and an upper bound ν^{Us}_{it}.

3. Uniformly sample ν^s_{it} from the interval $\left[\nu^{Ls}_{it},\nu^{Us}_{it}\right]$ for each participant to obtain a cost draw that is consistent with the observed behavior and the estimated parameters.

4. Compute equilibrium actions according to the following procedure:

   (a) For each participant $i = 1, ..., I_t$, choose a random starting action $d_{it}^s \in \{0, 1, ..., D\}$, where D is the submission limit.

   (b) Loop through participants, updating participant i's action according to

   $$d_{it}^s = \arg\max_{d_{it}} \left[R_t(d_{it}, d_{-it}^s; a_{it}, a_{-it}) - (\theta_1^s + \nu_{it}^s) d_{it} - \theta_2^s d_{it}^2 \right]$$

   for $d_{it} \in \{0, 1, ..., D\}$, where D is the counterfactual submission limit and $R_t(.)$ is the counterfactual contest expected returns function.

   (c) Repeat 4b until the updating procedure no longer changes participant actions. This rest-point is a Nash Equilibrium of the contest game.

5. Calculate contest outcome metric V_t^s at the equilibrium actions, the parameter vector θ^s and the cost draws $\{\nu_{it}^s\}_{i=1}^{I_t}$.

Steps 1-5 are repeated S times. In Table 6, I report the lower bound on the counterfactual outcome as $V_t^L = \min(V_t^s)$ and the upper bound as $V_t^U = \max(V_t^s)$. To obtain the average outcome across contests, as shown in Table 7, I use $V^L = \frac{1}{T}\sum\limits_{t=1}^{T} V_t^L$ for the lower bound and $V^U = \frac{1}{T}\sum\limits_{t=1}^{T} V_t^U$ for the upper bound.

A.4.2 Incomplete Information

The procedure described in Appendix A.4.1 can be modified as follows to incorporate incomplete information. First, in Step 2, all instances of $R_t(d_{it}, d_{-it}; a_{it}, a_{-it})$ must be replaced with $ER_{it}(d_{it})$, which can be obtained using the procedure described in Section 6.2. Then, the resulting cost intervals $\left[\nu_{it}^{Ls}, \nu_{it}^{Us}\right]$ will take into account that participants had incomplete information when choosing their actions. Second, Step 4 must be modified to capture the change in the equilibrium density of the number of competitors, participant actions and characteristics when the structure of the contest changes. Formally, Step 4 can be modified as follows, assuming that $\left[\nu_{it}^{Ls}, \nu_{it}^{Us}\right]$ have been obtained for all participants in the contests in $\mathcal{W}$ in a previous step.

4. Compute equilibrium actions according to the following procedure:

(a) For each participant $i = 1, ..., I_t$, set d_{it}^s to the participant's observed action.

(b) At iteration $k+1$, loop through participants and contests, updating participant i's action in contest t according to

$$d_{it}^{sk+1} = \arg\max_{d_{it}} \left[ER_{it}^k(d_{it}) - (\theta_1^s + \nu_{it}^s) d_{it} - \theta_2^s d_{it}^2 \right]$$

where

$$ER_{it}^k(d_{it}) = \frac{1}{B} \sum_{b=1}^{B} R_t(d_{it}, d_{-j_b b}^{sk}; a_{it}^b, a_{-j_b b})$$

for $d_{it} \in \{0, 1, ..., D\}$, where D is the counterfactual submission limit and $R_t(.)$ is the counterfactual contest expected returns function. Abilities a_{it}^b are constructed as in Section 6.2 and j_b denotes a random participant in contest b.

(c) Repeat 4b until $d_{it}^{sk+1} = d_{it}^{sk}$ for all $t \in \mathcal{W}$ and all i in contest t. This rest point is a Bayes-Nash equilibrium of the incomplete information contest game.

In general, the above procedure will only recover one of many possible equilibria. However, in simulations based on the complete information model, I find that if multiple equilibria exist at the same parameter vector, the behavior of participants across equilibria does not differ substantially and the implications for counterfactual outcomes are very similar.