Rewarding the Few or the Many? An Investigation of the Impact of Rewards in Open Innovation Contests

Reto Hofstetter, Pinar Yildirim, Z. John Zhang *

September 2013

Abstract

In this study, we examine the impact of financial incentives on participation and quality of submitted ideas in crowd-sourced innovation contests. We compare three commonly used contest designs that vary with respect to the number of winners and how the total reward is distributed: winner takes all (WTA), equally shared reward (ES, where n winners share a reward equally), and rank proportional rewards (RP, where n winners share a reward proportionally based on their ranking). We show that how the rewards are distributed amongst the winners will affect who participates in a contest and the quality of ideas submitted. We find that different financial incentive schemes are better for achieving different objectives. In particular, when the total reward is sufficiently large relative to the cost of participating in a contest, WTA is best at encouraging participation but not at generating ideas of the highest average quality in an open innovation platform. When the reward is small, the highest participation is achieved through RP. On the other hand, ES yields the highest average quality of submissions and the lowest participation when the reward is large. We use data from crowdSPRING.com to test our model predictions and our empirical findings are largely consistent with our theoretical conclusions.

Keywords: Open Innovation, Winner-Takes-All, Reward Structure, Participation, Crowdsourcing.

*Authors are listed alphabetically. All authors contributed equally. Hofstetter is Associate Professor of Marketing, University of Lugano, Via Giuseppe Buffi 13, CH-6904 Lugano. Email: reto.hofstetter@usi.ch, phone: +41 058-666-4860. Yildirim is Assistant Professor of Marketing, The Wharton School, University of Pennsylvania, Philadelphia PA 19104. E-mail: pyild@wharton.upenn.edu, phone: 215-746-2369. Zhang is the Murrel J. Ades Professor and Professor of Marketing, The Wharton School, University of Pennsylvania, Philadelphia PA 19104. E-mail: zjzhang@wharton.upenn.edu, phone: 215- 898-1989. We thank Wharton Mack Center for their support and Remo Siegwart for supporting data collection. Also thanks to Christophe Van den Bulte, Josh Eliashberg and the participants of the Marketing Science 2013 and EMAC 2011 conferences for their helpful comments.

1 Introduction

Innovation is the engine of growth, and not surprisingly, firms pay close attention to open innovation contests. The number of open innovation platforms and contestants has been on the rise in the past decade. For example, CrowdSPRING.com, the leading open innovation contest platform, boasts a network of more than 130,000 creators from 185-plus countries (crowdspring.com). Another platform, Innocentive, recently disclosed that it has registered over 270,000 problem-solvers from nearly 200 countries and awarded over $37 million to solvers in 1500 challenges to date (innocentive.com).

The increasing popularity of open innovation contests is fueled by seekers (e.g., firms) that repeatedly engage the same *pool* of participants in a platform for new ideas. On CrowdSPRING, 30% of all seekers hosted more than one contest on the platform in 2011, and the total number of contests is growing at a 9% annual rate. General Motors, Harley Davidson, and other major firms are now generating their advertising ideas through crowd-sourcing. Other firms, like Starbucks or DELL, host their own permanent platforms where customers are invited to post their ideas. These firms understand that hosting contests can offer several benefits along the innovation funnel, such as access to diverse solutions at low cost and the ability to screen the top ideas out of a vast solution space (Terwiesch and Ulrich (2009)).

As firms use crowdsourcing platforms, they naturally face the questions of how best to incentivize crowd participation and attract highest quality ideas. In this paper we investigate these questions both theoretically and empirically for low budget crowdsourcing contests.

In practice, various reward structures, such as the winner-takes-all (WTA), equally shared rewards for top participants (ES), and rank proportional rewards (RP) for top participants are used by innovation platforms. However, it is not clear what type of incentives allow seekers to achieve their participation and quality goals. For seekers interested in collecting as many ideas as possible, achieving a high participation level in a contest can be an important goal. For example, a seeker may want to solicit advertising ideas for its product marketing campaigns and through the process it may also want to engage its customers. To this end, many companies now crowd-source their SuperBowl commercials as a way to engage their base. Other seekers may care less about participation but more about the quality of ideas submitted by participants. For example, a seeker may try to generate the design for a brand's logo and attract few but the highest quality participants. In either case, a firm must use the right incentive structure to accomplish its objective. The question is: which one? In this study, we investigate what the right incentive structure is for accomplishing the desired objective by focusing on contests that require *minimum effort* or *expertise* and are frequently conducted in open innovation platforms.

In our theoretical model, each experienced contestant is drawn from a pool of contestants who care about their probability of winning and the size of the prize if they win, and who weigh their expected financial return against their cost of entry under a reward structure. Each experienced participant is assumed to have had private feedback on the quality of her previous submission, which we assume correlated perfectly with her ability. This private information allows the participant to update her probability of winning before entering a contest. At the equilibrium, all participants behave similarly and therefore the equilibrium participation and average quality of submitted ideas can be calculated.

The distinguishing features of this study from the existing ones in open innovation literature are three-fold. First, we consider open innovation contests as a game between multiple participants in response to reward incentives. Second, we investigate the impact of three commonly used financial incentives in contests to uncover how reward schemes influence the outcomes of participation and the quality of submitted ideas in a contest. Third, we test our model implications using data from crowdSPRING.com.

We find that each incentive scheme can be optimal depending on (i) whether a seeker is after more participants vs. higher average quality [1] of submitted ideas, and (ii) the

[1]We focus on the average quality simply because the quality of a single idea is a probabilistic event and a firm can only count on achieving the highest average quality when repeatedly hosting crowdsourcing contests.

attractiveness of the reward relative to the cost of a participant's entry. When the goal is participation, a high reward to cost ratio implies that WTA is preferable, while RP is more preferable if the reward to cost ratio is low. However, when obtaining a high average quality of ideas is the goal, ES is preferred with a high reward and WTA is preferred with a low reward. This is because, as we uncover, each reward scheme displays a unique level of efficiency, for a given size of the total reward for a contest, in delivering a prize to the winner(s) and achieves different winning probabilities as the total reward size differs. Together, the winning prize and the winning probabilities will incentivize participants of varying qualities to self-select and join a contest, yielding different levels of participation and average quality of submitted ideas.

Three forces play a role in determining which incentive is more preferable in open innovation contests. The first relates to the positive externalities from the enhanced winning probability in *multi*-winner contests. We refer to this effect as the "everybody is a winner effect". Participation is higher due to the attractiveness of multiple winning spots when the total reward is distributed among n winners (equally or proportional to rank) compared to when it is awarded to a single participant. However, when the total reward is shared among multiple winners, a second and counter effect rises: the per person reward increases at a slower rate for the highest quality participants in multi-winner contests. We refer to this effect as the "trophy effect". Finally, because the "everybody is a winner" and the "trophy" effects attract contestants at different rates, the probability of winning changes at different rates as other contestants enter the contest. We refer to this effect as the "crowding out effect", which reduces participation for multiple winner incentives at a higher rate than WTA incentives when the total reward is high. Combining the three effects, as the total reward increases, distributing it amongst n winners *equally* becomes the least efficient way of allocating the prize money to winners. We find that participation under ES incentive is lower than the participation under RP or WTA incentives. However, this lower participation with ES incentive has the effect of inducing self-selection: only high quality participants choose to enter a contest, as they have a higher probability of winning resulting in the highest average quality submissions with ES.

While our study focuses on the theoretical incentive design for running small open innovation contests, we are also motivated by the commercial success of such contests in an industrial context. After analyzing what each incentive can and cannot do, we also seek to confront our conclusions with a large-scale dataset from crowdSPRING.com containing information on 16,038 contests. We exploit the variation in reward schemes on the site to test the predictions of our analytical model about the nature of the relationships between the reward scheme and the outcome variables of interest. The empirical results indicate that the outcomes on crowdSPRING are indeed in line with most of our predictions. Most notably, we find that participation is the highest with RP if the total reward is low and with WTA if the total reward is high. The highest average contest quality is achieved by choosing WTA or RP reward scheme when the total reward is low and an ES or RP reward scheme if the total reward is high.

Our study contributes to the existing literature in three ways. First, we add to the rich literature on innovation by focusing on the behavior of a group of agents instead of a single agent or a single firm (e.g., Schmittlein and Mahajan (1982), Hauser et al. (2006)), as the presence of multiple agents is a significant characteristic of open innovation contests. Second, past academic investigations on open innovation contests have focused mostly on the design of contests from a perspective of (i) how many participants to allow in a single contest, (ii) how the number of participants influence the number of ideas and effort, and (iii) how the incentive schemes and rewards influence the effort level. Terwiesch and Xu (2008) develop a theoretical model to investigate the type of problems most suited for innovation contests with a high stake (i.e., a large total reward), and identify the optimal reward and the optimal number of solvers to reach out to. Different from this paper, we focus on the contests that require minimum effort and little expertise, and build a model of experienced participants where a key driving force is each participant's anticipation of other participants' entry into a contest. Erat and Krishnan (2012) focus on the total number of ideas developed by solvers for design problems. Boudreau et al. (2011) study data from Topcoder, a contest platform that is for elite software developers. Although this environment would be very suitable for investigating repeated participation in contests, their investigation is limited to analysis at a single point in time and focuses on how the number of participants affects the participants' effort level. Our study, unlike Boudreau et al. (2011), focuses on contests that are open to amateurs and require little expertise and effort. Bayus (2012) generates an empirical analysis of individual's ideation efforts in an open innovation contest using publicly available data from Dell's Ideastorm Community. He tests empirically how ideas an individual generated in the past influence his future idea generation.

More closely related to our study is Archak and Sundararajan (2009). The paper mostly focuses on designing the incentive scheme for an open innovation contest depending on whether participants are risk-neutral or risk-averse. They conclude that when participants are risk-neutral, the seeker should allocate its entire contest budget to the top winner, even if he values multiple submissions. When participants are sufficiently risk averse, the seeker should distribute the prize budget across more prizes than the number of submissions he desires. In our study, participants choose to enter based not only on incentives offered in a contest, but also on how many other contestants they anticipate will enter the same contest. In addition, we examine how to choose three commonly used reward schemes for different reward budgets by investigating the efficacy of these three schemes in achieving two specific seeker goals: participation and high average quality. Finally, our study also adds to the growing literature on user-generated content (UGC). There have been various studies to date on UGC (e.g., Yildirim et al. (2013), Shriver et al. (2013)) but few focus on the innovation aspect of user-driven creation in marketing, either from a theoretical or empirical perspective. The involvement of users in open innovation contests nicely expands the UGC literature.

The rest of the article is organized as follows. In Section 2, we develop our model of open innovation contests. In Section 3 we generate several insights into which incentive scheme works best under various seeker goals. In Section 4 we test these findings by analyzing data collected from crowdSPRING.com. Finally, we provide a discussion and conclude with managerial implications and limitations of our work.

2 The Model

We built our model with two aims in mind. First, we wanted to develop a model that captures the unique characteristics of open innovation contests as closely as possible. Second, we wanted to understand the economics behind how different incentives motivate participants in a *simple* model to allow for clean empirical testing using our crowdSpring dataset. All the pieces we abstracted away from modeling, we control for them in the empirical section (see Section 4).

Open innovation contests differ from other mechanisms of innovation in several ways. First, there are clear pre-announced financial incentives to submitting an idea. Second, seekers seek solutions from users of products who often have low expertise and professional experience in conducting the task.[2] In addition, due to the large pool of contestants, the repeated interaction among agents is very low, preventing a single participant from gathering perfect information about others. Finally, the tasks are often of a creative nature (e.g., logo design, narratives) and awards are frequently moderate to matter for effort. For those reasons, we abstract away from modeling effort and learning.

Our modeling section will follow two stages. First, we demonstrate how the *ex ante* pool of contestants is affected when agents have imperfect knowledge of their own and other contestants' quality. Second, we model *experienced* contestants making entry decisions after they receive feedback on their own quality.

[2]This seems to be the case particularly for our platform of interest, crowdSpring.com. Similarly, Starbucks and Dell, for example, use the ideas generated by consumers in their open innovation platforms to improve their products.

2.1 Formation of Initial Pools of Contestants

Consider a platform holding innovation contests for a seeker. To define a pool of contestants for the platform, we start with a situation where contestants participate in a specific type of contest for the first time. Given that a contestant has no information about her own as well as other contestants' skills, she will enter a contest hosted by the platform if her expected reward is non-negative. Formally, let A be the total reward presumably determined through a budgetary process outside of our model, and let a_i, $i = 1, 2, ...n$ be the per person reward for the top n winners. A contestant incurs the fixed entry cost f when she enters a contest. This cost, for example, can refer to the opportunity cost of working on a submission. A rational contestant will enter the contest if

$$Prob_i(\text{winning})a_i - f \geq 0, \tag{1}$$

where $Prob_i(\text{winning})$ is the probability that contestant i will win the prize a_i. The winning probability in equation (1) will depend on the incentive schemes a seeker uses.

In the case of the "Winner Takes All (WTA)" incentive scheme, only one winner is selected for each contest. The seeker awards all A to the single top winner so that $a_i = A$ and $Prob_i(\text{winning}) = \frac{1}{N}$, where N is the expected number of contestants in the contest. The entry and exit by contestants will continue until equation (1) holds in equality, so that we have $N = A/f$.

In the case of "Equally Shared" incentive scheme (ES), a seeker distributes A equally amongst the top n winners, with $a_i = \frac{A}{n} > f$. The probability of winning in this case is $\frac{n}{N}$. Therefore, similar to the previous case, in equilibrium we must have $N = A/f$.

Another common incentive scheme is to award top n participants as in the previous case but divide the total reward A based on the rank (quality) of them. For simplicity, we assume that the reward allocated to participant i holding rank k in a contest is:

$$a_{ik} = \frac{2A(n + 1 - k)}{(n + 1)n}, k = 1, 2, .., n. \tag{2}$$

This means, for example, if a total reward of \$15 is allocated amongst 5 winners, the participant with the 1^{st} rank would obtain \$5, the participant ranked 2^{nd} would obtain \$4, the participants ranked 3^{rd} would obtain \$3, and so on. In this case, given a priori identical contestants, the expected reward for a contestant is given by:

$$\sum_{k=1}^{n} \frac{2A(n+1-k)}{(n+1)n} \frac{1}{N} = \frac{A}{N}. \tag{3}$$

In equilibrium, we have again $N = A/f$ contestants in the initial pool. Thus under all three incentive schemes, the size of a contest's reward and cost of entry determine the size of the pool of contestants and they are identical. This feature captures the long term determinants of the size of the contestants pool. It also gives us the advantage of conducting our analysis by holding the size of the contestants pool *constant across different incentive schemes*.

As a seeker frequently taps the same pool of participants more than once and offers feedback on the quality of a contestant's submission, an experienced participant out of this pool can update her beliefs about probability of winning a prize based on the feedback she receives. We assume that she has imperfect information about the quality of other contestants. However, it is common knowledge that the quality of each contestant's idea (q) is a random draw with replacement from a uniform distribution, i.e., $q \in U[0, 1]$. This assumption, as mentioned earlier, captures the unique features of open innovation contests: first, a contestant typically has *little information* about other participants, and second, the pool of the contestants is generally *large* and the cohort of people participating in the same consecutive contests is generally *small* [3]. As a result, each participant will evaluate her own quality against the distribution of the quality in the same pool of participants.

[3] These assumptions are justified in the statistics from our crowdSPRING data, as demonstrated in Section 4.

2.2 Experienced Contestant Participation

We now investigate how the experienced contestants make entry decisions and how these decisions affect the average quality of ideas submitted under each incentive scheme.

2.2.1 Winner Takes All (WTA)

An experienced contestant with private signal $q = q_W \in U[0,1]$, who expects $\tilde{N}_W$ contestants in the contest, will enter a WTA contest if she expects the quality of every other entry to be lower than hers. This happens with a probability of ${q_W}^{(\tilde{N}_W - 1)}$. This means the marginal consumer $\bar{q}_W$, who is indifferent between participating and not participating, is determined by:

$$\bar{q}_W^{(\tilde{N}_W - 1)} A - f = 0. \tag{4}$$

In equilibrium, the expected number of participants in the contest should be confirmed, so that we also have $\tilde{N}_W = (1 - \bar{q}_W)N = (1 - \bar{q}_W)A/f$. Therefore the two equations that define the equilibrium number of participants ($\tilde{N}_W$) and the marginal consumer ($\bar{q}_W$) are given by:

$$\begin{aligned} ((1 - \bar{q}_W)A/f - 1)\, ln(\bar{q}_W) &= ln(\frac{f}{A}), \\ \tilde{N}_W &= (1 - \bar{q}_W)A/f. \end{aligned} \tag{5}$$

We measure the average quality of ideas submitted in the contest by calculating the expected quality conditional on entering the contest. For a known $\bar{q}_W$, participants share the unit quality dq as:

$$\int_{\bar{q}_W}^{1} \frac{q}{(1 - \bar{q}_W)}\, dq = \left(\frac{1 + \bar{q}_W}{2}\right), \tag{6}$$

where we focus on the average quality simply because the quality of a single idea is a probabilistic event and a firm can only count on achieving the highest average quality when repeatedly hosting crowdsourcing contests.

2.2.2 Equally Shared Rewards (ES)

Let's denote the quality of the marginal contestant under ES with $\bar{q}_{ES}$ and the equilibrium number of participants with $\tilde{N}_{ES}$. In that case, the analogous two equations that define the equilibrium under ES are given by:

$$\begin{aligned} ((1-\bar{q}_{ES})A/f-n)ln(\bar{q}_{ES}) =& ln(\frac{nf}{A}), \\ \tilde{N}_{ES} =& (1-\bar{q}_{ES})A/f. \end{aligned} \tag{7}$$

The average quality of ideas is obtained by the conditional expectation $\frac{(1+\bar{q}_{ES})}{2}$ derived similarly to equation (6).

2.2.3 Rank Proportional Reward (RP)

After receiving her private feedback q, a contestant updates her likelihood of being at each of top n ranks as follows. When a contestant is ranked 1^{st}, it implies that her quality is higher than the quality of the remaining $N-1$ contestants in the contest, which happens with a probability of q^{N-1}. Similarly, if the contestant is ranked 2^{nd}, this implies that her quality is higher than $(N-2)$ of the contestants but lower than one of the $(n-1)$ contestants rated in the top n. Then the probability of being ranked 2^{nd} is $q^{N-2}(1-q)\binom{n-1}{1}$. The probability of attaining each of the top n ranks is then given as:

$$\text{Probability of Ranks} = \begin{cases} q^{N-1} & 1^{st} Rank \\ q^{N-2}(1-q)^1\binom{n-1}{1} & 2^{nd} Rank \\ q^{N-3}(1-q)^2\binom{n-1}{2} & 3^{rd} Rank \\ q^{N-4}(1-q)^3\binom{n-1}{3} & 4^{th} Rank \\ . \\ . \\ q^{N-n}(1-q)^{n-1}\binom{n-1}{n-1} & n^{th} Rank. \end{cases} \tag{8}$$

The sum of the probabilities given in equation (8) amounts to q^{N-n}, same as in the probability of winning in ES incentive. In addition, if $n = 1$, equation (8) is reduced to q^{N-1}, the probability of winning under WTA as expected.

A contestant will enter an RP contest if her expected reward satisfies:

$$\sum_{k=1}^{n} q^{N-k}(1-q)^{k-1}\binom{n-1}{k-1}\frac{2A(n+1-k)}{(n+1)n} - f \geq 0. \tag{9}$$

Let $\bar{q}_{RP}$ denote the quality of the marginal contestants and $\tilde{N}_{RP}$ denote the equilibrium number of participants under RP. The two equations that determine the equilibrium are given by:

$$\frac{2A}{(n+1)n}\sum_{k=1}^{n} \bar{q}_{RP}^{\tilde{N}_{RP}-k}(1-\bar{q}_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k) - f = 0, \tag{10}$$

$$\tilde{N}_{RP} = (1-\bar{q}_{RP})\frac{A}{f},$$

Similar to the previous two incentive schemes, in equilibrium the average quality is calculated using $\frac{(1+\bar{q}_{RP})}{2}$.

3 Comparison of the Incentive Schemes

In this section, we compare the three incentive schemes for degree of participation by experienced contestants and the average quality of ideas submitted in such a contest. As the initial contestant pools are identical for all three incentive schemes, any difference in participation levels and average quality can be attributed solely to the fact that different incentive schemes motivate contestants differently through the probability of winning and the awards for the winners. Proposition 1 summarizes our findings with regard to participation. The proofs for all the propositions in this section are provided in the Appendix.

Proposition 1. *If the total reward (A) is small relative to the entry cost (f), RP is the most effective incentive scheme at inducing contestants' participation in an innovation contest. Otherwise, WTA is the most effective.*

Intuitively, an experienced contestant will weigh two interrelated factors in determining whether to continue her participation: the reward if she wins ("the trophy effect") and the probability of winning. The probability of winning will depend, under each incentive scheme, on the number of rewards (n) given in a contest ("everybody is a winner effect" or EIAW effect in short) and also the number of other participating contestants (N) which in turn depends on the size of the total reward (A) ("crowding out effect"). The conclusions in Proposition 1 reflect the fact that different incentive schemes will affect these factors differently to result in different participation rates.

In the case of WTA, the winning probability for any contestant is always low as there is only one winner, and decreases slightly with the total reward size as the bigger reward size induces more contestants to enter, and hence dilutes a contestant's chance to win via the "crowding out effect". For WTA the "EIAW effect" is absent but the size of the winning reward increases monotonically and dollar for dollar with the size of reward. Therefore, the expected reward decreases at a lower rate for WTA as the total reward becomes larger. Said differently, in WTA, the "trophy effect" is large (see Figures 1 and 2 for examples of expected reward for two agents with $q = 0.8$ and $q = 0.85$ and Figure 3 for how probability of winning changes with total reward for each incentive). Combining the three effects, WTA is the *most efficient* in delivering a contest's reward to the winner, but the *least efficient* in delivering the winning probability. Consequently, it is most effective at motivating participation when the "trophy effect" overshadows the "crowding out" effect, which happens if the total reward in a contest is large. Interestingly, we commonly observe this outcome in the case of Powerball lotteries: a high prize for a single winner induces more entries into lotteries.

[INSERT FIGURES 1, 2 and 3 HERE]

With the RP incentive scheme, the total reward is shared by n winners so that the reward per winner increases much more slowly than in the case of WTA. The "trophy effect" is thus weaker with RP. However, the EIAW effect is present and a contestant has n opportunities to win the contest. The probability of winning a prize with RP is always higher relative to that of WTA, especially when the total number of contestants is small (see Figure 3). Remember that the EIAW effect suggests that contests with multiple winners are generally more attractive to participants compared to contests with only one winner. However, multiple winning spots come at the cost of having to share the total award with other contestants. For that reason, only when the total reward is small the RP incentive scheme stands out in motivating participation because of its high winning probability. In other words, RP motivates participation chiefly through the "EIAW effect" that increases the probability of winning, but this effect is fast diminishing as the number of contestants increases with the total reward in a contest, or as the "crowding out effect" becomes more dominant. WTA, on the other hand, becomes more attractive due to faster increasing per person reward.

The case of ES is intermediate between WTA and RP incentive schemes: it is less efficient in delivering rewards to the winners (or the “trophy effect” is weaker) especially to high quality winners relative to RP, but it is more efficient in generating a high probability of winning (or the “EIAW effect” is stronger) relative to WTA. This combination of efficiencies will not allow this incentive scheme to stand out in motivating participation (as shown in Figure 5), but it has the advantage of delivering the right incentives to high quality contestants as we will see in the next proposition.

Proposition 2. *If the total reward (A) is small relative to the entry cost (f), WTA is the most effective incentive scheme at generating higher average quality submissions in an innovation contest. Otherwise, ES is the most effective one.*

Under ES, the participation rate is the lowest with a higher reward because the reward per person under ES is insensitive to the relative ranks of the winners. With respect to the total reward, ES is the incentive scheme favored by low quality participants as they have a higher per person reward compared to RP. RP is favored by high quality participants as they have much higher probability of making it to the top ranks and be rewarded more generously once they get there. As the rewards in a contest always go to the top quality contestants, ES is the least effective way to allocate any prize to high quality winners. This means that a contestant with q sufficiently high, she foresees less expected reward with ES incentive than in the RP case for the same total reward A (see Figure 4). For that reason, one would need to have a higher q to enter in the ES case than in the RP case. In other words, the inefficiency in delivering the rewards to the winners allows ES to screen out low quality contestants and to select the higher quality ones (see Figure 5 for how the contest ‘participating segment’ changes with respect to A).

[INSERT FIGURES 4 and 5 HERE]

Propositions 1 and 2 together suggest that how the prize is meted out is of great importance in running crowdsourcing contests and depending on a firm's main objective and budget, the firm should choose different incentive schemes as we summarize in Proposition 3.

Proposition 3. *The choice of an incentive scheme in running a crowdsourcing contest can help a firm to achieve a high participation level or high average quality objective and this choice should depend on the size of the total reward offered for the contest. In particular, as shown in Figure 6,*
(i) when the total reward is small, RP yields the highest participation and WTA yields the highest average quality of ideas,
(ii) if the reward size is large, WTA yields the highest participation and ES yields the highest average quality of ideas.
(iii) Otherwise, when the reward size is medium, RP yields the highest participation and ES yields the highest average quality of ideas.

To provide a numeric example on how participation and and average quality of ideas move with respect to the total reward, we fixed the entry fee (f) and the number of winners (n). Figure 6 illustrates this schedule when $n = 10$ and $f = 1$. For each total reward (A), the highest participation level or highest average quality is achieved by different financial incentives. For example, when A is low (i.e., less than roughly 80 in the figure), a seeker aiming for high continued participation is better off holding a contest with RP incentive, and a seeker who aims for higher quality ideas should hold a contest with WTA incentive. When A is medium, ($80 < A < 126$ in the figure), a seeker with a participation objective should use RP and a seeker aiming for high quality ideas should use ES. Finally, when distributing a large reward ($A > 126$ in the figure), the seeker should use WTA to attract more contestants, and distribute the reward equally among 10 contestants if aiming for higher quality ideas. Suffice it to say, our research suggests that different reward sizes and seeker objectives call for different incentive mechanisms, which is the underlying message of our article.

Of course, which incentive scheme yields the highest overall participation and the highest average quality depend on the financial attractiveness of an innovation contest A. As noted in Figure 6, when $n = 10$, increasing reward A increases the participation levels and average quality for all incentive schemes. This result should be shown algebraically and we formalize it in the following proposition.

Proposition 4. *For all incentive schemes, as the total reward (A) increases, the total number of participants and average quality of submissions increase.*

[INSERT TABLE 1 AND FIGURE 6 HERE]

We provide a summary of the discussion from the current section and the best incentive for each seeker's objective in Table 1. We also summarize the testable hypotheses stemming from the analytical model here:

1. When the total reward A increases (Proposition 4)

   H_{1a} : total number of participants increases,

   H_{1b} : average quality of submitted ideas increases.

2. Number of participants is (Proposition 1)

   H_{2a} : the highest under the WTA incentive when the total reward (A) is high,

   H_{2b} : the highest under the RP incentive when total reward (A) is low,

   H_{2c} : always higher under RP compared to under ES incentive.

3. Average quality of ideas is (Proposition 2)

   H_{3a} : the highest under the ES incentive when total reward (A) is high,

H_{3b} : the highest under the WTA incentive when A is low,

H_{3c} : always higher under ES compared to under RP.

4 Empirical Evidence from crowdSPRING.com

4.1 Data

We now turn to empirical evidence to see if the main conclusions of our model have external validity. For this purpose, we collected a large-scale dataset from crowdSPRING.com, one of the leading open innovation platforms and marketplaces for buyers and sellers of creative services (Russ (2011)). We crawled all publicly available information of the website. The resulting data includes all contests dating back to the very first contest ever held on the platform on April 1^{st} in 2008. After its launch, the site went through a nine week beta testing phase during which developers worked on its functionality and later, adjusted the reward system. After June 20th, they changed the feedback policy and encouraged seekers to provide feedback to the problem solvers during the contest (see blog.crowdspring.com). As our model conclusions are based on equilibrium analysis, we shall only consider contest data after this early and unstable period of the site.

The resulting dataset contains information on $16,038$ contests held by $11,767$ seekers among $36,654$ solvers. Seekers who ran more than a single contest hosted a substantial proportion of the contests (40.92%). Counting multiple contest observations per seeker allows us to treat the data as panel information, suitable for our empirical analysis, as discussed in the next section. We also repeatedly observe solvers as $24,490(66.81\%)$ participated in at least one additional contest. In total, we count $526,114$ contest participations and $1,727,386$ posted submissions. The median seeker hosts one contest and gets 34 participants and 112 submissions per contest on average. The median solver participated in 3 contests and made 7 submissions in total. Overall, $4,550(12.41\%)$ solvers won at least one reward and the winners on average received $1,339.60 over all of their contest participations.

Contests are held in 39 different design and writing subcategories. The most prominent of these are creative tasks such as logo design, logo and stationary design, print design, and web-design, in the order of decreasing prominence (see Table 2 for the number of contests held in each subcategory). In each contest, a seeker determines the total reward to be awarded, the number of winners, and how the reward is distributed amongst the winners. All three incentive schemes discussed in Section 2 are observed during the data period, making crowdSPRING.com an ideal place to test the propositions summarized in Section 3.

[INSERT TABLE 2 HERE]

A descriptive analysis of the data collected reveals that the WTA incentive is predominantly favored by seekers and was used in $15,373 (95.85\%)$ of all contests.[4] RP is chosen in $340 (2.12\%)$ contests and ES is chosen in $325 (2.03\%)$ contests (a summary of other contest statistics is provided in Table 3.) Even though RP and ES incentives have been utilized far fewer times compared to WTA incentive, the number of observations for both is sufficiently large to make meaningful statistical inferences.

[INSERT TABLE 3 HERE]

4.2 Contest Participation and Quality Models

We start the empirical analysis by investigating the effect of total award A and the incentive scheme on (i) the number of participants and (ii) the mean quality of the submissions in a contest. We first build our main model and then address further complications that rise from estimation in the following sections.

[4]This popularity might be partially due to WTA being the default incentive for a contest on the platform. This skewness in the data is taken into account in the analysis.

We define a system of two structural equations as follows. For each contest i that was initiated by seeker j, our data consist of observations for the total reward A_{ij}, the number of entrants N_{ij}, the mean quality of a contest q_{ij}^{avg}, as well as the choice of the incentive scheme, RP, ES, or WTA. In equations (11) and (12), N_{ij} and q_{ij}^{avg} are dependent on the incentive scheme and a set of covariates X_{ij}.

$$N_{ij} = \gamma_{10} + \gamma_{11} WTA_{ij} + \gamma_{12} ES_{ij} + \gamma_{13} A_{ij} + \gamma_{14} WTA_{ij} \times A_{ij} + \gamma_{15} X_{ij} + \varepsilon_1, \quad (11)$$

$$q_{ij}^{avg} = \gamma_{20} + \gamma_{21} WTA_{ij} + \gamma_{22} ES_{ij} + \gamma_{23} A_{ij} + \gamma_{24} WTA_{ij} \times A_{ij} + \gamma_{25} X_{ij} + \varepsilon_2. \quad (12)$$

The first column in Table 4 summarizes the necessary conditions that allow us to find support for the predictions given in Section 3. We are interested in the relationship between total reward and the number of contestants (captured by γ_{13}) and the quality of submissions (captured by γ_{23}). To allow the intercepts to vary between WTA, ES and RP contests, we include coefficients γ_{11}, γ_{12}, γ_{21}, γ_{22} for each incentive scheme with RP incentive serving as the baseline. Since RP and ES incentives are qualitatively similar (they are both multiple winner incentives), we include the interaction terms between WTA and A_i only (captured by γ_{14} and γ_{24}) to allow WTA to differ in slope from that of ES and RP. This structure allows testing the hypotheses in Section 3.

[INSERT TABLE 4 HERE]

4.3 Complications and Extensions

In this section, we address four complications specific to our data as well as to open innovation contests in general: Common unobservables driving both contest participation and quality of ideas, different rating styles of seekers, missing seeker feedback, and correlation between seeker's feedback and participation levels. We propose that these issues can be resolved by standardizing the dependent variables and by including specific control variables for feedback and participation levels. We discuss how the empirical model can be adjusted accordingly below and further test the impact of each of the discussed complications on our estimates in the empirical section.

Common Unobservables. First, there may be unobservables driving both the number of entrants and quality. For instance, a seeker's brand reputation may attract a higher number and higher quality of solvers to a contest. In such a scenario, the error terms in the equations will be positively correlated ($cov(\varepsilon_1, \varepsilon_2) = \sigma_{12}, \sigma_{12} \neq 0$). Although OLS can separately estimate these two equations, efficiency gains could be achieved if we use GLS to account for the correlated unobservables. Alternatively, we can explicitely control at least for time-constant seeker related unobservables because of the specific panel nature of our data. Similar to the idea of time-demeaning in fixed effects estimation, mean-centering the dependent variables quantity drops these unobservables from the equations. We favor this option since it simplifies the estimation task. If a seeker j hosts a total number of K_j contests, then the mean centered measure for contest quantity $\dot{N}_{ij}$ of N_{ij} is:

$$\dot{N}_{ij} = N_{ij} - \frac{\sum_{i=1}^{K_j} N_{ij}}{K_j} = N_{ij} - \bar{N}_j. \tag{13}$$

Notice what happens to the error term of (11) after mean centering the dependent variable. If we assume an unobserved time-constant seeker popularity effect p_j, it would be captured by the error term in the initial equation:

$$N_{ij} = \ldots + \varepsilon_1, \varepsilon_1 = p_j + u_{ij}. \tag{14}$$

After mean centering, p_j drops from the error term if we assume that $p_j = \bar{N}_j$.

$$\dot{N}_{ij} = N_{ij} - \bar{N}_j = \ldots + \varepsilon_1, \varepsilon_1 = p_j - \bar{N}_j + u_{ij}. \tag{15}$$

The same logic applies to our measure of mean contest quality.

Seeker's Rating Style. On crowdSPRING.com, each seeker rates submissions it receives on a scale from 1 to 5, where 5 indicates the highest quality. Each seeker might have its specific style of evaluating the submissions that could result in a shift of mean quality. Stricter seekers consistently provide lower ratings while friendlier seekers rate higher. Panel data proves to be beneficial also in this case as the previously discussed mean-centering will remove such systematic shifts. In addition to mean-centering, the quality measure needs to be normalized by dividing the mean-centered values by the standard deviation sd_{ij} of the ratings as seekers could also differ in the rating scales they apply. Some might use a narrow scale while others use a wide range of the scale. Hence, we apply the following standardization to the quality measure:

$$\dot{q}_{ij}^{avg} = \frac{q_{ij}^{avg} - \frac{\sum_{i=1}^{K_j} q_{ij}}{K_j}}{sd_{ij}}. \tag{16}$$

We adjust the two equations by replacing the dependent variables by their standardized versions. Please notice that we also remove the intercept as a consequence of the standardization:

$$\dot{N}_{ij} = \gamma_{11} WTA_{ij} + \gamma_{12} ES_{ij} + \gamma_{13} A_{ij} + \gamma_{14} WTA_{ij} \times A_{ij} + \gamma_{15} X_{ij} + \varepsilon_1, \tag{17}$$

$$\dot{q}_{ij}^{avg} = \gamma_{21} WTA_{ij} + \gamma_{22} ES_{ij} + \gamma_{23} A_{ij} + \gamma_{24} WTA_{ij} \times A_{ij} + \gamma_{25} X_{ij} + \varepsilon_2. \tag{18}$$

Missing Seeker Feedback. The third complication we address is related to missing data. We observe that on crowdSPRING.com, seekers often fail to provide feedback for all the submissions they receive.

Researchers can explain these data gaps in two ways. In the first and less problematic case, a seeker has evaluated part of the submissions, which consists of 91.79% of the contests in our data. In this case, we can derive the mean quality with some measurement error whenever a systematic rating style is present. For instance, in order to minimize effort, a seeker may target a subset of submissions it received in its evaluative efforts such as focusing only on high quality submissions. In the event of such selective evaluation behavior, one would expect to see high ratings present in contests with low number of submissions. We find evidence in the data to support this explanation, as mean rating in a contest is negatively correlated with the percentage of submissions rated ($corr = -.44, p = .000$). This behavior might both be seeker and contest specific. Through the standardization of the dependent variable we already controlled for the time-constant rating style of the seeker but we did not control for a potential variation in rating style over time. We thus add a control variable that accounts for the percentage of submissions rated (*prcrated*) to equation (18).

In the second and more problematic case, a seeker has not evaluated any of the submissions. Calculation of a mean rating for such contests is not possible. In order to avoid subjecting our results to a 'selection bias' (Heckman (1976)), we prefer not to ignore the contests in which no submissions are rated although they account for a small portion of our data (7.25% of all contests). In this case, the coefficients can still be consistently estimated with full maximum likelihood or Heckman's two-step estimator (see Greene and Zhang (1997)). In Section 4.4, we discuss how we deal with this second scenario. As we uncover, this potential source of bias appears to be systematic but does not substantially influence our results. We thus conclude that the factors that lead to missingness are mostly unrelated to the reward system and we thus believe it is safe to ignore such missing data for our analysis.

Participation and Seeker Feedback. A fourth source of complication may stem from a seeker's tendency to give a limited number of high ratings to winners. As only one or a few solvers can win a contest, the top ratings provided by the seeker will be limited. That is, only a few top submissions would get a top rating of 5 stars and the mean quality measure will be downward biased with higher numbers of entrants. Figure 7 illustrates this concern. Whereas the densities of high and low ratings are balanced for the contests with low numbers of entrants (left graph), lower ratings are more likely in contests with a high numbers of entrants (right graph).

[INSERT FIGURE 7 HERE]

We thus expect a negative effect of $\dot{N}_{ij}$ on $\dot{q}_{ij}$ as a result. To control for this third bias, we add $\dot{N}_{ij}$ to equation (18) which subsequently reads as follows:

$$\dot{q}_{ij}^{avg} = \beta_{21}\dot{N}_{ij} + \beta_{22}prcrated_{ij} + \gamma_{21}WTA_{ij} + \gamma_{22}ES_{ij} + \gamma_{23}A_{ij} + \gamma_{24}WTA_{ij} \times A_{ij} + \gamma_{25}X_{ij} + \varepsilon_2. \tag{19}$$

An estimation via OLS using (19) is potentially biased since $\dot{N}_{ij}$ is the dependent variable in equation (11) and we assumed $cov(\varepsilon_1, \varepsilon_2) \neq 0$. To obtain an unbiased estimate, we include an instrumental variable which impacts $\dot{N}_{ij}$ but is unrelated to unobservables that drive $\dot{q}_{ij}^{avg}$. We test for such a correlation of the error terms and check whether an instrumental variable approach would improve estimation in the results section.

4.4 Separate Estimation Results

In this section, we lay out the details of the estimation of equations (17) and (19). We start by separately estimating the equations followed by a discussion of the potential gains with a simultaneous estimation. We find that the separate estimation is most appropriate in our case as that mean-centering the dependent variables already controls for potentially correlated unobservables and inclusion of $\dot{N}_{ij}$ does not add substantial endogeneity concerns. We thus favor the results of the separate estimation (Model 3 in Table 5 for the participation equation and Model 4 in Table 5 for the quality equation) and perform all hypotheses testing using these estimates.

[INSERT TABLE 5 HERE]

Participation Equation. We start with Model 1 (see Table 5) where we estimate the coefficients for total reward A and incentive schemes with no control variables added. The model yields a significant positive coefficient for A and a significant negative effect of ES but no significant effects for WTA and WTA × A [5]. Next, in Model 2, we enhance Model 1 by adding both *category* and *year* controls. Adding these dummies is crucial because both the choice of total reward amount (A) and incentive scheme vary among categories due to the different opportunity costs of participation. In addition, the year dummies can correct for any time trends or temporal variations in participants' choices. Omitting both variables can potentially result in an 'omitted variables' bias. Indeed, the effect of A is reduced, the effect of WTA is also reduced and becomes significant, and WTA $\times A$ flips sign. These controls explain a substantial proportion of the variance as adjusted R^2 is substantially increased from 0.007 to 0.122 in Model 2.

Model 3 adds *contest* and *seeker* specific controls. In our theoretical model, we assume these variables are fixed, therefore we need to control for them here. First, we control for tenure of a seeker using the variable *contestcount*, which represents the number of contests a seeker has previously hosted on crowdSPRING. We expect that a seeker that hosted a higher number of contests previously may attract more solvers in subsequent contests possibly due to reputation. Second, we control for the duration of the contest by adding the variable *duration* which indicates the number of days a contest lasts. The longer a contest lasts, the more participants we expect to have in the contest. Third, *prcrated* is the previously discussed measure of the percentage of submissions the seeker evaluates, or the seeker's diligence level. We control for this factor as well. This variable is bounded between 0 and 1 and continuous.

Interestingly, the coefficient of *prcrated* is negative indicating that fewer participants enter a contest if seekers are more diligent in rating participants. It could be that such seekers tend to be more critical and thus less attractive for solvers. Other explanations

[5] Please note that we measure A in \$1,000 so an effect of 6.929 indicates increasing A by 1,000 resulting in an absolute increase of 6.929 in the number of participants.

are possible here and we do not take a strong stance on the exact process that led to this result.

Finally, the variables *featured* and *assured* stand for two contest features that can be chosen by a seeker when setting up a contest. First, a featured contest ($featured = 1$) is more prominently displayed on crowdSPRING.com, and hence more seekers will become aware of the contest and might choose to participate. The coefficient for this variable is positive, indicating that hosting a featured contest results in a 1.969 increase in the number of contestants. Second, a seeker can provide assurance in advance that some reward will be handed out regardless of the quality of submissions ($assured = 1$). An even larger increase of 2.733 participants can be achieved if the seeker provides this assurance. Adding these controls slightly reduces the coefficient of A, an expected outcome because they correlate slightly and positively with A.

The results indicate that the coefficients of the model are in line with the predictions of the analytical model as outlined in Section 3. We observe a positive association between the total reward and the number of participants for RP and ES contests. The intercept for ES is lower for RP and even lower for WTA contests. The relationship with A is stronger for WTA contests. Taken together, these factors form the three regions given in Figure 6, implying that for low levels of A, RP is associated with the highest number of participants and for high levels of A, WTA is associated with the highest number of participants, as Proposition 1 in the analytical model concluded. The effects are plotted out in Figure 8, which shows the out-of-sample predictions[6]. The plot visualizes the three regions predicted by Proposition 3 in the analytical model. The first intercept is at position $A_1 = \frac{\gamma_{12}-\gamma_{11}}{\gamma_{14}}$ and the second at $A_2 = \frac{-\gamma_{11}}{\gamma_{14}}$. The left part of the region cannot fully be confirmed as A_1 is not significantly different from zero (A_1=.278, t=1.10, p=.273). The middle part, however, is significant as both A_2 (A_2=1.476, t=3.62, p=.000) and $A_2 - A_1$ (($A_2 - A_1$)=1.198, t=2.24, p=.025) are significant. Based on the model estimates we can

[6] A varies between 0 (\$0) and 2 (\$2000), all other continuous variables are held at their mean and categorical variables at their mode.

also test the previously outlined hypotheses. The results of these tests are summarized in Table 4. We can confirm H_{1a}, H_{2a}, H_{2b}, and H_{2c}.

Quality Equation. Estimation of equation 19 is more complicated because of inclusion of the dependent variable of the first equation in the second equation and the potential measurement error in the dependent variable. The estimation results are summarized in Table 6.

[INSERT TABLE 6 HERE]

In Table 6, Model 1 shows the results when only incentive dummies and an intercept term are included, whereas Model 2 also controls for contest *category* and *year* effects. We see that adjusted R^2 is 0.027 even after including these controls. This indicates a much lower ability to explain variance in the quality equation compared to the participation equation. The coefficient of the total reward (A) is significant, however the intercept of ES is not. In Model 3, we add further controls including two potential sources of measurement bias: the number of entrants ($\dot{N}_{ij}$) and percentage of rated submissions (*prcrated*). Both variables are highly significant and negatively correlated with mean quality, providing support for the expectation of measurement bias in *reported* quality data. Moreover, the coefficient of A is positive and significant but significantly lower for WTA contests, in line with what we would expect to see according to the theoretical model.

In Model 4, we also add the tests for the impact of missing data (i.e., lack of feedback information). Ignoring the missing cases can bias the estimates if the process that has led to the missingness is not completely random. Further, it is possible that the process that leads to missingness also influences the mean quality feedback in contests. For instance, if the submissions are of very low quality, the seeker may feel less motivated to provide feedback to participants biasing the quality measure in the positive direction.

We utilize Heckman's (Heckman (1976)) two-step estimator to test interdependence of the quality and the selection equation. The selection equation is estimated first and the nonselection hazard $\lambda(x'\beta)$ (or inverse mills ratio) is derived and then included in the main equation. The hypothesis of independence between the selection and main equation can be directly tested using λ, which represents the covariance $\sigma_{sel,main}$ of the error terms of the two models. We estimate the Heckman model without explicit exclusion restrictions using the same covariates of the main model in the probit model and identify the parameters through the non-linear functional form of the effects of the covariates (see Cameron and Trivedi (2005)).

We find the selection model explains 17.39% of the variance. Selection is mainly explained by contest characteristics that are independent of the reward system. Suffice it to say, λ in Heckman's two-step estimator is significantly different from zero (λ=.056, z=2.05 p=.041), which suggests dependence of the two equations. However, we do not find strong differences between the estimates of the reward scheme coefficients between Models 3 and 4. As a result, we conclude that the factors associating with a seeker failing to evaluate any submissions are partly the factors driving the mean quality feedback, but the reward scheme covariates are not dramatically affected by this interdependence.
We use the estimates of Model 4 for our hypothesis testing. Again, we first visualize the effects by graphing out the out-of-sample predictions, which are depicted in Figure 8 (right graph). Once more too, we observe the structure as hypothesized in the theoretical section.

For the quality equation, the first intercept is at position $A_1 = \frac{\gamma_{22}-\gamma_{21}}{\gamma_{24}}$ and the second at $A_2 = \frac{-\gamma_{21}}{\gamma_{24}}$. The left part of the region is confirmed as A_1 is significantly different from zero (A_1=.825, t=2.89, p=.004). The middle part is relatively narrow and thus not significant ((($A_2 - A_1$)=.232, t=.66, p=.512)) but the right part is again confirmed (A_2=.593, z=2.25, p=.025). Hypothesis H_{1a} can only be confirmed for RP and ES contests but not for WTA contests. H_{3a} and H_{3b} can also only partly be confirmed and H_{3c} is not confirmed. Overall, however, we conclude that also for the quality equation, the results are largely in line with our predictions. While some hypotheses cannot explicitly be confirmed, we also find no strong evidence that would lead to a rejection.

4.5 Simultaneous Estimation and Results

We next consider the potential need for a simultaneous estimation of the equations (17) and (19) that could arise because of the inclusion of $\dot{N}_{ij}$ into the quality equation and common unobservables that might remain after mean-centering of the dependent variables.

We argued that $\dot{N}_{ij}$ could be endogenous in equation (19) when $cov(\varepsilon_1, \varepsilon_2) \neq 0$. Estimation can still be consistent with an instrument that drives $\dot{N}_{ij}$ but is unrelated to $\dot{q}_{ij}^{avg}$. Fortunately, the panel nature of our data can be exploited to derive such an instrument. Specifically, we use the temporal correlation between the number of entrants to the current contest and the number of entrants to the previous contest of a given seeker. Such temporal lags are often used as instrumentals (see e.g. Stahl et al. (2012), Vilcassim et al. (1999), Lambrecht and Misra (2013)) because it can often conveniently be assumed that the temporal correlation only occurs within a given variable of interest and not between variables. While the number of entrants may affect the quality measure in the current contest, the number of entrants in past contests is unlikely to influence the quality measure in the current contest (i.e., $cov(\dot{N}_{ijt-1}, \dot{q}_{ijt}^{avg}) = 0$ at any time t) after controlling for $\dot{N}_{ijt}$. We believe this assumption is plausible because a seeker's contests do not share many similarities apart from the time-constant unobservables as we find that the solver pool is subject to high temporal variation. Solvers are not loyal to specific seekers and the structure of the crowd almost fully changes between contests of the same seeker: solvers participate in only .42 % of all observed contests of the same seeker immediately following a contest, and only in 2.28% of the contests from the same seeker overall. As a result, only .03% of the participants in a contest have previously competed against another solver from a previous contest of the same seeker. Hence, the seeker is confronted with an almost entirely different set of solvers for each contest, suggesting that time-varying unobservables driving both $\dot{N}_{ijt-1}$ and $\dot{q}_{ijt}^{avg}$ may not be present. Note that the time-constant seeker unobservables are already controlled for through the use of time-demeaning or centralization of the measures.

We can empirically confirm a significant effect of the lag on the number of entrants (see Model 1 in Table 7). Even though one cannot test the assumption of no correlation with unobservables, we show that the instrument has no predictive power on mean quality, after controlling for the covariates in the model (as given by the t-values in parentheses):

$$\dot{q}_{ij}^{avg} = -.007(-6.71) \times N_t + -.000(-0.08) \times N_{t-1} + \ldots controls \ldots + \varepsilon_2. \qquad (20)$$

Subsequently, we apply two-stage least squares to estimate equation (19). The results are given in Table 7 under 'Model 1'. We test whether the instrument we used is working properly by combining the LM and Wald versions of the rk statistic from Kleibergen-Paap (2006). The underidentification hypothesis is strongly rejected (rkLM = 98.135, p =0.000) and the F-statistic is with 188.078 strongly above the critical value suggested by Stock and Yogo (2005) of 16.38. We do not need to test overidentifying restrictions as our equation is just identified. We thus conclude that the instrument is working as intended.

[INSERT TABLE 7 HERE]

Even though the instrument appears to work, we do not find a significant change in the coefficients after using it, according to the results of the insignificant endogeneity test (F(1,2021) = 0.007, p = 0.934)[7].

To investigate the potential impact of common unobservables, we explicitly estimate the covariance of the error terms σ_{ij} and jointly estimate the two equations with full information maximum likelihood (FIML), a typically more efficient estimator in situations of interdependence. The results are shown in Model 2 of Table 7. Surprisingly, although

[7] We report the Wooldridge (1995) robust score test hence since we ran two-stage least squares with errors clustered at the seeker level.

some of the standard errors are lower in accordance to the higher efficiency of the joint estimation (please note that all standard errors are clustered at the seeker level for Model 3 and bootstrapped for Model 4), the estimates do not change. Most interestingly, σ_{12} is insignificant, indicating the equations' independence. The two equations can thus be estimated separately. It appears that the unobserved heterogeneity not already captured by the time-constant seeker's popularity and rating style is small. This result is plausible and suggests that the mean-centering of the dependent variables largely removes unobserved heterogeneity simplifying the estimation procedure.

5 Conclusion, Implications, and Limitations

In this study, we investigate the financial incentives offered in small stake open innovation contests and their impact on participation and the quality of ideas submitted in a contest. We develop an analytical framework to compare three commonly used financial incentives and we then test our key propositions empirically using data collected from crowdSPRING.com. To uncover the impact of incentives, we recognize that seekers in open innovation contests care about different performance metrics. For example, while some seekers might be interested in generating high participation, others may focus on high average quality of ideas. Therefore, from the perspective of a seeker interested in launching an open innovation contest, it is equally important to ask which incentive is linked to the highest number of participation and the highest quality submissions.

Our theoretical findings show that three effects determine which incentive scheme yields the highest participation and highest average quality of ideas. First, the schemes that allow multiple participants to win are more effective in attracting participation than WTA, *ceteris paribus* ("everybody is a winner effect"). However, each incentive scheme rewards high quality participants differently ("trophy effect"). For each additional dollar allocated to a contest, WTA incentive allows the winner to be awarded at the highest rate (i.e., dollar for dollar) whereas RP and ES incentives are less efficient in rewarding top participants. Finally, as the total reward increases, the increase in the number of participants have the adverse effect of discouraging participants from entry by diminishing the probability of winning ("crowding out effect").

We show that which effect plays a dominant role depends on the incentive and the value of the total reward. When the total reward A is low, the trophy effect plays a relatively less important role in attracting participants to a contest. In addition, EIAW effect is absent with WTA. As a result, WTA performs poorly in attracting participants compared to the other two incentives with a low A. On the other hand, as the total reward increases, trophy effect and crowding out effect together suppress EIAW to allow WTA to be the most efficient incentive in attracting participants and ES to be the least efficient. The reduced appeal of the ES incentive at high A level has, nevertheless, the desired effect of inducing participants' self-selection: only high quality participants choose to enter a contest. Increased self-selection raises the quality of submissions: only highest quality contestants will choose to enter under ES at high A.

In our empirical analysis, we refrain from making causal claims but rather investigate the relationships between the incentives, participation, and average quality of ideas as predicted by the analytical model. The results largely support the propositions derived analytically. For participation, the results essentially mirror the predictions of the model. We find that if a relatively low total reward is offered on crowdSPRING, RP contests are associated with the highest number of participants, followed by ES, and then WTA incentives. With high total rewards, the number of participants grows relatively more strongly for WTA compared to RP and ES. As a result, WTA incentive yields the highest participation when total reward is relatively high.

The empirical results for average quality are also largely supportive of the predictions of our analytical model. For RP and ES, we find a positive significant effect of A, and for WTA the direction is positive, albeit not significant. We believe that the partial empirical support for average quality predictions stems from the noise in the seeker ratings. In particular, the ratings for contests with WTA incentive have low reliability, since the seekers have lower motivation to rate every submission.

We intended to provide a simple analysis of the most commonly used open innovation contest incentive schemes. Our study can guide managers who run open innovation contests as we highlight the importance of (1) total reward and (2) incentive scheme selection as determinants of (1) participation and (2) innovative quality a manager can expect from such contests. Specifically, a manager interested in customer participation (average quality of ideas) should choose WTA (ES) if her contest budget is large, and RP (WTA) if it is small.

However, there are some limitations to the current work. We consider only three incentive schemes commonly used in the industry, and it is possible that there are other ones which are not considered here. We focus on two goals for a seeker, and there may be other objectives than the ones considered in this paper. In addition, our results apply to contests where effort is not a significant factor in determining the winning probability. This assumption captures the reality of the small stake contests we study. Future research could extend this study to consider how incentives affects effort and in turn determines quality. Finally, the careful reader may consider whether the selection of the prize money correlates with the dependent variables. We believe that the total reward allocated is mostly a function of firm budgetary constraints of the nature of contests.

Appendix

Proof of Propositions 1 and 2:

To prove Proposition 3, we will compare the number of participants $\frac{A(1-q)}{f}$ under each incentive scheme for a given point of A and f. We also make the same comparisons based on the average quality $\frac{(1+q)}{2}$. For ease of notation, we will drop the upperbars from each q term in the proof. The q for the marginal participant under each incentive scheme is defined by:

$$Aq_W^{(N_W-1)} = f, \tag{A.1}$$

$$q_{ES}^{(N_{ES}-n)}\frac{A}{n} = f, \tag{A.2}$$

$$\frac{2A}{(n+1)n}\sum_{k=1}^{n} q_{RP}^{N_{RP}-k}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k) = f. \tag{A.3}$$

As we have

$$\frac{2A}{(n+1)n}\sum_{k=1}^{n} q_{RP}^{N_{RP}-k}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k)$$
$$=\frac{2A}{(n+1)n}q_{RP}^{N_{RP}-n}\sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k),$$

we re-write equation (A.3) as:

$$\frac{2\zeta}{(n+1)}q_{RP}^{(N^{RP}-n)}\frac{A}{n} = f, \tag{A.4}$$

where $\zeta \equiv \sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k)$.

To prove Proposition 3, we first show that $q_{ES} > q_{RP}$ for all $q_{RP} > \frac{1}{2}$. We then show that q_W intersects first with q_{ES} and then with q_{RP} as A increases to generate the three regions discussed in Proposition 3.

To show that $q_{ES} > q_{RP}$ for all $q_{RP} > \frac{1}{2}$, first note that $q_{RP} = q_{ES}$ if $\frac{2}{(n+1)}\zeta = 1$ holds based on equations (A.2) and (A.4). If $\zeta > \frac{n+1}{2}$ or $\frac{2}{(n+1)}\zeta > 1$, then we must have $q_{RP} < q_{ES}$. We now proceed to show that indeed $\zeta > \frac{n+1}{2}$.

Note we must have either $(n+1-2k) > 0$ or $(n+1-2k) < 0$, except at $k = \frac{n+1}{2}$. In the case of $(n+1-2k) > 0$, given $q_{RP} > \frac{1}{2}$ and hence $\left(\frac{q_{RP}}{1-q_{RP}}\right) > 1$, we must have

$$\left(\frac{q_{RP}}{1-q_{RP}}\right)^{(n+1-2k)} > 1.$$

This must then imply

$$\left(\frac{n+1}{2}-k\right)\left(\frac{q_{RP}}{1-q_{RP}}\right)^{(n+1-2k)} > \left(\frac{n+1}{2}-k\right). \tag{A.5}$$

In the case of $(n+1-2k)<0$, then we have

$$\left(\frac{q_{RP}}{1-q_{RP}}\right)^{(n+1-2k)} < 1.$$

Multiplying both sides by the negative term $(\frac{n+1}{2}-k)$, equation (A.5) still holds. Simple manipulation of the inequality in (A.5) yields

$$\left(\frac{n+1}{2}-k\right) q_{RP}^{(n-k)}(1-q_{RP})^{k-1} > \left(\frac{n+1}{2}-k\right)\left(q_{RP}^{(k-1)}(1-q_{RP})^{n-k}\right), \text{ or}$$

$$\begin{aligned}(n+1-k)q_{RP}^{(n-k)}(1-q_{RP})^{k-1} &+ kq_{RP}^{(k-1)}(1-q_{RP})^{n-k} \\ &> \frac{n+1}{2}\left(q_{RP}^{(n-k)}(1-q_{RP})^{k-1}+q_{RP}^{(k-1)}(1-q_{RP})^{n-k}\right).\end{aligned} \tag{A.6}$$

Notice that the summation of the LHS of the inequality from $k=1$ to $k=\frac{n}{2}$ for some even number n yields

$$\begin{aligned}&\sum_{k=1}^{k=\frac{n}{2}}\left((n+1-k)q_{RP}^{(n-k)}(1-q_{RP})^{k-1}+kq_{RP}^{(k-1)}(1-q_{RP})^{n-k}\right),\\ &=\sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k)=\zeta.\end{aligned}$$

The first equality holds because the second term $k\,q_{RP}^{(k-1)}(1-q_{RP})^{(n-k)}$ summed from $k=1$ to $k=\frac{n}{2}$ is equal to the summation of the first term $(n+1-k)q_{RP}^{(n-k)}(1-q_{RP})^{k-1}$ from $k=\frac{n}{2}+1$ to $k=n$. Similarly, summing up the terms on the RHS of equation (A.6) yields

$$\sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}\frac{(n+1)}{2}.$$

Therefore, we must have

$$\zeta = \sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k) > \sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}\frac{(n+1)}{2}.$$

Clearly this inequality holds even if $(n+1-2k)=0$ for a single term. Using this inequality, we set the lower bound on the parameter ζ as follows:

$$\begin{aligned}\zeta =& \sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1}(n+1-k)\\ >& \frac{n+1}{2}\sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{k-1}\binom{n-1}{k-1} = \frac{n+1}{2}(q_{RP}+1-q_{RP})^{(n-1)} = \frac{n+1}{2},\end{aligned}$$

where due to Binomial Theorem

$$\sum_{k=1}^{n} q_{RP}^{(n-k)}(1-q_{RP})^{(k-1)}\binom{n-1}{k-1} = (q_{RP}+1-q_{RP})^{(n-1)} = 1.$$

Therefore, as $\zeta > \frac{(n+1)}{2}$ we can conclude that $q_{RP} < q_{ES}$ is true for any $q_{RP} > \frac{1}{2}$. *Q.E.D.*

We next compare the total reward at which the WTA and ES incentives result in the same marginal consumer. Let there be a total reward A^* at which point the marginal contestant under ES and WTA incentives are identical, $q^* = q_W^* = q_{ES}^*$ and the expected return from WTA and ES incentives are aligned as follows:

$$q_W^{*\,(N_W-1)}A^* = q_{ES}^{*\,(N_{ES}-n)}\frac{A^*}{n} = f. \tag{A.7}$$

Rewriting this equation,

$$n(q_{ES})^{(\frac{A^*}{f}(1-q_{ES})-1)} = q_{ES}^{(\frac{A^*}{f}(1-q_{ES})-n)}, \tag{A.8}$$

or equivalently,

$$n(q)^{*(n-1)} = 1, \text{ or } q^* = \left(\frac{1}{n}\right)^{\frac{1}{n-1}}. \tag{A.9}$$

This closed form solution implies that $\exists q^*, s.t. \frac{1}{2} \leq q^* = q_W^* = q_{ES}^* < 1$ for each $n \geq 2$, under the assumption that the number of entrants is higher than the number of awards given, or $\frac{A(1-q^*)}{n} > f$ holds.

Next, let's compare the WTA and RP incentives. Similar to before, let there be a total reward A^{**} at which point the marginal participants under WTA and ES incentives are identical $q^{**} = q_W^{**} = q_{RP}^{**}$ and the expected return from the two contests are aligned as follows:

$$q_W^{**(N_W-1)} A^{**} = \frac{2\zeta A^{**}}{n(n+1)} q_{RP}^{**(N_{RP}-n)} = f. \tag{A.10}$$

where $N_W = \frac{A^{**}}{f}(1 - q_W^{**}) = N_{RP} = \frac{A^{**}}{f}(1 - q_{RP}^{**})$. Simplifying this expression implies that at some A^{**}:

$$\frac{n(n+1)}{2\zeta} q^{**^{(n-1)}} = 1. \tag{A.11}$$

Since ζ is a function of q^{**} (i.e., $\zeta(q^{**}) = \sum_{k=1}^{n} q^{**(n-k)}(1-q^{**})^{k-1}\binom{n-1}{k-1}(n+1-k)$), equation (A.11) can be rewritten as

$$\frac{n(n+1)}{2\left(\sum_{k=1}^{n}\left(\frac{1-q^{**}}{q^{**}}\right)^{k-1}\binom{n-1}{k-1}(n+1-k)\right)} = 1. \tag{A.12}$$

We now show that the q^{**} satisfying equation (A.12) must lie between 0 and 1 by contradiction.

At $q^{**} = \frac{1}{2}$, the LHS of equation (A.12) is less than 1 since:

$$\begin{aligned}\frac{n(n+1)}{2\left(\sum_{k=1}^{n}\left(\frac{1-\frac{1}{2}}{\frac{1}{2}}\right)^{k-1}\binom{n-1}{k-1}(n+1-k)\right)} =& \frac{n(n+1)}{2\sum_{k=1}^{n}\binom{n-1}{k-1}(n+1-k)}\\ <& \frac{n(n+1)}{2\left(n+(n-1)(n-1)\right)}\\ <& 1.\end{aligned}$$

On the other hand, when $q^{**} \to 1$, LHS reaches and exceeds 1, and approaches to infinity. Thus, there must exist a solution q^{**} that satisfies equation (A.12) such that $1 > q^{**} > \frac{1}{2}$.

Next, we proceed to show that the solution to equation (A.12) is unique. Because the denominator of the term on LHS is monotonically decreasing with q^{**} for any $1 > q^{**} > 0$, the whole term on the LHS is monotonically increasing with q^{**}. This implies that there can only be one q^{**} which satisfies equation (A.12).

Finally, we show $q^{**} > q^{*}$. From equation (A.11) we see that if $\zeta = \frac{n+1}{2}$, we must have $q^{**} = q^{*}$. However, as shown before we have $\zeta > \frac{n+1}{2}$ so that $\frac{n(n+1)}{2\zeta}$ evaluated at q^{**} is smaller than the expression evaluated at q^{*}. Thus q^{**} must be larger than q^{*} to maintain the equality.

Thus on the plane of participation rate $1-q$ and total reward A, there are two regions separated by A^{**}, below which RP dominates and above which WTA dominates, as illustrated in Figure A.1. The total number of participants ($\frac{A(1-q)}{f}$ are illustrated in Figure A.2 This proves Proposition 1.

On the plane of quality q and total reward A, there are two regions separated by A^{*}, below which WTA dominates and above which ES dominates, as illustrated in Figure A.3. This proves Proposition 2. *Q.E.D.*

□

Figure A.1: Total Reward vs. Participation Rate for RP, ES, and WTA.

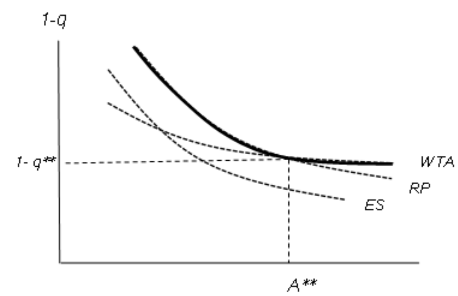


Figure A.2: Total Reward vs. Total Participation for RP, ES, and WTA.

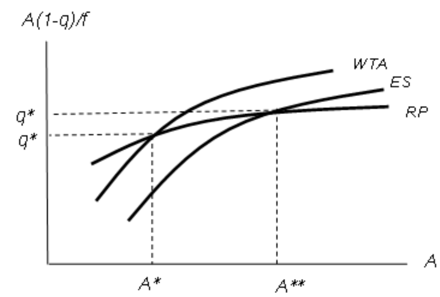
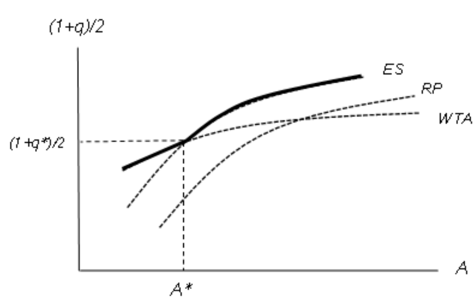


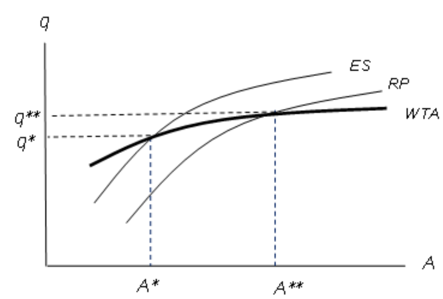
Figure A.3: Total Reward vs. Average Submission Quality for RP, ES, and WTA.



Proof of Proposition 3:

To prove Proposition 3, recall that $q_{ES} > q_{RP}$ and $q^{**} > q^{*}$ holds. Given $\frac{\partial q_w}{\partial A} > 0$ (which is demonstrated in the proof of Proposition 4), then we must have the total reward corresponding to q^{**} and q^{*} satisfying $A^{**} > A^{*}$. Therefore, on the plane of q and A we must have three regions as summarized below and Figure A.4 demonstrates these three regions.

Figure A.4: Total Reward vs. Marginal Contestant Quality for RP, ES, and WTA.



- if $A \in (A_0, A^{*}]$ then $q_W \geq q_{ES} > q_{RP}$,
- if $A \in (A^{*}, A^{**}]$ then $q_{ES} > q_W \geq q_{RP}$,
- if $A \in (A^{**}, \infty)$ then $q_{ES} > q_{RP} > q_W$,

where $A_0 > \frac{fn}{1-q_{RP}}$ is defined as the lowest limit for A below which there will be more winners than participants. *Q.E.D.* $\square$

Proof of Proposition 4:

Part i: $\frac{\partial}{\partial A}\frac{A(1-q)}{f} > 0$**:**

Let's demonstrate the relationship for WTA incentive. The derivative can be expressed as:

$$\frac{\partial A(1-q_W)/f}{\partial A} = \frac{1}{f}((1-q_W) - A\frac{\partial q_W}{\partial A}).$$

Since $\frac{1}{f} > 0$, we are interested in the sign of the remaining term $((1-q_W) - A\frac{\partial q_W}{\partial A})$. Remember that the participation for WTA equation is:

$$q_W^{N_W-1}A - f = 0.$$

Let's define the function F as

$$F \equiv (N_W - 1)log(q_W) - log(f/A) = \left(\frac{A(1-q_W)}{f} - 1\right) log(q_W) - log(f/A) = 0.$$

With implicit derivation, the sign of the derivative can be obtained with the relationship $sign(\frac{\partial q_W}{\partial A}) = sign(-\frac{\frac{\partial F}{\partial A}}{\frac{\partial F}{\partial q_W}})$.

$$\frac{\partial F}{\partial q_W} = \frac{\partial[(\frac{A(1-q_W)}{f} - 1)log(q_W)]}{\partial q_W} + \frac{\partial log(f/A)}{\partial q_W} = -\frac{A}{f}log(q_W) + \frac{\frac{A(1-q_W)}{f} - 1}{q_W}. \quad \text{(A.13)}$$

In addition,

$$\frac{\partial F}{\partial A} = (\frac{\partial N^W}{\partial A})log(q_W) + \frac{1}{A} = \frac{(1-q_W)}{f}log(q_W) + \frac{1}{A}. \quad \text{(A.14)}$$

Combining the two terms given in equations (A.13) and (A.14),

$$\frac{\partial q_W}{\partial A} = (-\frac{\frac{\partial F}{\partial A}}{\frac{\partial F}{\partial q_W}}) = -\frac{\frac{(1-q_W)}{f}log(q_W) + \frac{1}{A}}{-\frac{A}{f}log(q_W) + \frac{\frac{A(1-q_W)}{f} - 1}{q_W}}.$$

And

$$
\begin{aligned}
(1-q_W) - A\frac{\partial q_W}{\partial A} &= (1-q_W) + \frac{\frac{A(1-q_W)}{f}log(q_W)+1}{-\frac{A}{f}log(q_W)+\frac{\frac{A(1-q_W)}{f}-1}{q_W}} \\
&= \frac{(1-q_W)\left(-\frac{A}{f}log(q_W)+\frac{\frac{A(1-q_W)}{f}-1}{q_W}\right)+\frac{A(1-q_W)}{f}log(q_W)+1}{-\frac{A}{f}log(q_W)+\frac{\frac{A(1-q_W)}{f}-1}{q_W}} \\
&= \frac{(1-q_W)\left(\frac{\frac{A(1-q_W)}{f}-1}{q_W}\right)+1}{-\frac{A}{f}log(q_W)+\frac{\frac{A(1-q_W)}{f}-1}{q_W}} \\
&> 0.
\end{aligned}
$$

Note that $log(q_W) < 0$ and $\frac{A(1-q_W)}{f} - 1 > 0$, thus both the numerator and the denominator is positive, and the overall term is also positive.

Part ii: $\frac{\partial q}{\partial A} > 0$:

The sign of the derivative $\frac{\partial q}{\partial A}$ determines how participation rate and quality changes with respect to the total reward A. We now determine the sign of this derivative. For WTA:

$$
q_W^{N^W-1}A - f = 0.
$$

Let's define the function F as

$$
F \equiv (N^W-1)log(q_W) - log(f/A) = \left(\frac{A(1-q_W)}{f} - 1\right)log(q_W) - log(f/A) = 0. \tag{A.15}
$$

With implicit derivation, the sign of the derivative can be obtained with the relationship

$sign(\frac{\partial q_W}{\partial A}) = sign(-\frac{\frac{\partial F}{\partial A}}{\frac{\partial F}{\partial q_W}})$. $\frac{\partial F}{\partial q_W} > 0$ since

$$\frac{\partial F}{\partial q_W} = \frac{\partial[(\frac{A(1-q_W)}{f} - 1)log(q_W)]}{\partial q_W} + \frac{\partial log(f/A)}{\partial q_W} = \underbrace{-\frac{A}{f}log(q_W)}_{>0} + \underbrace{\frac{\frac{A(1-q_W)}{f} - 1}{q_W}}_{>0}. \quad (A.16)$$

In addition,

$$\frac{\partial F}{\partial A} = (\frac{\partial N^W}{\partial A})log(q_W) + \frac{1}{A} = \frac{A(1-q_W)}{f}log(q_W) + 1. \quad (A.17)$$

Let's express the solution q_W using $m \equiv N^W - 1 = \frac{A(1-q)}{f} - 1 > 0$.

$$q_W = (\frac{f}{A})^{\frac{1}{m}}.$$

Note that the Taylor series expansion for the logarithm of number $t < 1$ is:

$$log(t) \approx (t-1) - \frac{(t-1)^2}{2} + \frac{(t-1)^3}{3} \cdots$$

Since $t < 1$, each term is negative in this approximation. Therefore we can set an upper bound for the term: $log(t) < (t-1)$. Similarly, using Taylor series expansion, a bound on the $log(q_W)$ term can be set:

$$log(q_W) = \frac{1}{m}log(\frac{f}{A}) < \frac{1}{m}(\frac{f}{A} - 1).$$

Let's use this upper bound in equation (A.17)

$$\begin{aligned}\frac{\partial F}{\partial A} = \frac{A(1-q)}{f}log(q)+1 &< \frac{A(1-q)}{f}\frac{1}{m}(\frac{f}{A}-1)+1 \\ &= (1-q)\frac{1}{m} - \frac{A(1-q)}{f}\frac{1}{m}+1 \\ &= \frac{1}{m}((1-q) - \frac{A(1-q)}{f} + m) \\ &= \frac{1}{m}((1-q) - \frac{A(1-q)}{f} + \frac{A(1-q)}{f} - 1) \\ &= \frac{1}{m}((1-q)-1) \\ &= \frac{1}{m}(-q) \\ &< 0.\end{aligned}$$

Therefore $\frac{\partial F}{\partial q_W} < 0$. Given $sign(\frac{\partial q_W}{\partial A}) = sign(-\frac{\frac{\partial F}{\partial A}}{\frac{\partial F}{\partial q_W}})$ implies $\frac{\partial q_W}{\partial A} > 0$. (The proofs for $\frac{\partial q_{ES}}{\partial A} > 0$ and $\frac{\partial q_{RP}}{\partial A} > 0$ follow a similar route.) This proves Proposition 4. *Q.E.D.*

□

References

Archak, N., A. Sundararajan. 2009. Optimal design of crowdsourcing contests. *ICIS*. 200.

Bayus, Barry. 2012. Crowdsourcing new product ideas over time: An analysis of the dell ideastorm community. *Management Science* **Forthcoming** 606–623.

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

Cameron, A Colin, Pravin K Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Erat, S., V. Krishnan. 2012. Managing delegated search over design spaces. *Management Science* **58**(3) 606–623.

Greene, W.H., C. Zhang. 1997. *Econometric analysis*, vol. 3. Prentice hall Upper Saddle River, NJ.

Hauser, J., G. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing Science* **25**(6) 687717.

Heckman, J.J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models 475–492.

Lambrecht, Anja, Kanishka Misra. 2013. Pricing online content: Fee or free? *Working Paper* .

Russ, L. 2011. 100,000 designers and writers now work on crowdspring. *Crowdspring.com* .

Schmittlein, David C., Vijay Mahajan. 1982. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science* **1**(1) 57–78.

Shriver, Scott K, Harikesh S Nair, Reto Hofstetter. 2013. Social ties and user-generated content: Evidence from an online social network. *Management Science* **59**(6) 1425–1443.

Stahl, Florian, Mark Heitmann, Donald R Lehmann, Scott A Neslin. 2012. The impact of brand equity on customer acquisition, retention, and profit margin. *Journal of Marketing* **76**(4) 44–63.

Terwiesch, C., K. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business School Press, Boston, Massachusetts.

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

Vilcassim, Naufel J, Vrinda Kadiyali, Pradeep K Chintagunta. 1999. Investigating dynamic multifirm market interactions in price and advertising. *Management Science* **45**(4) 499–518.

Wooldridge, Jeffrey M. 1995. Score diagnostics for linear models estimated by two stage least squares. *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao* 66–87.

Yildirim, Pinar, Esther Gal-Or, Tansev Geylani. 2013. User-generated content and bias in news media. *Management Science* **59**(12) 2655–2666.

Figure 1: Total Reward (A) vs. Expected Reward ($q = 0.8, n = 10,\ f = 1$)

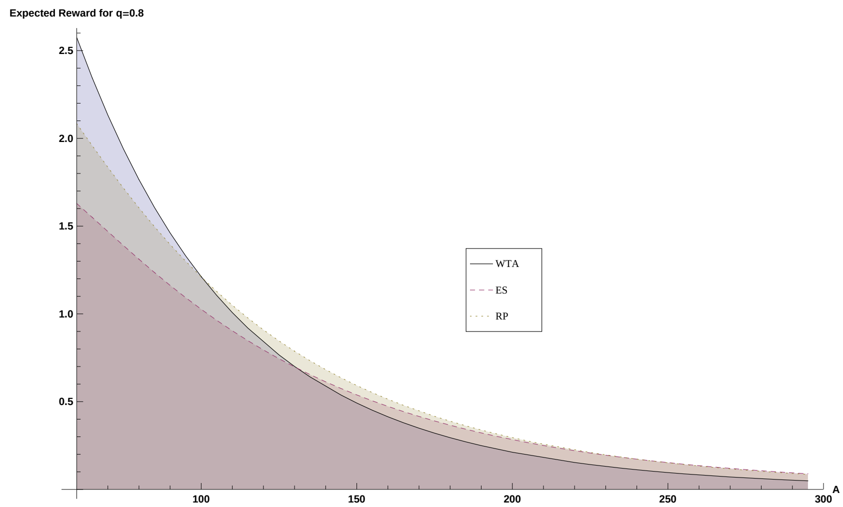


Figure 2: Total Reward (A) vs. Expected Reward ($q = 0.85, n = 10,\ f = 1$)

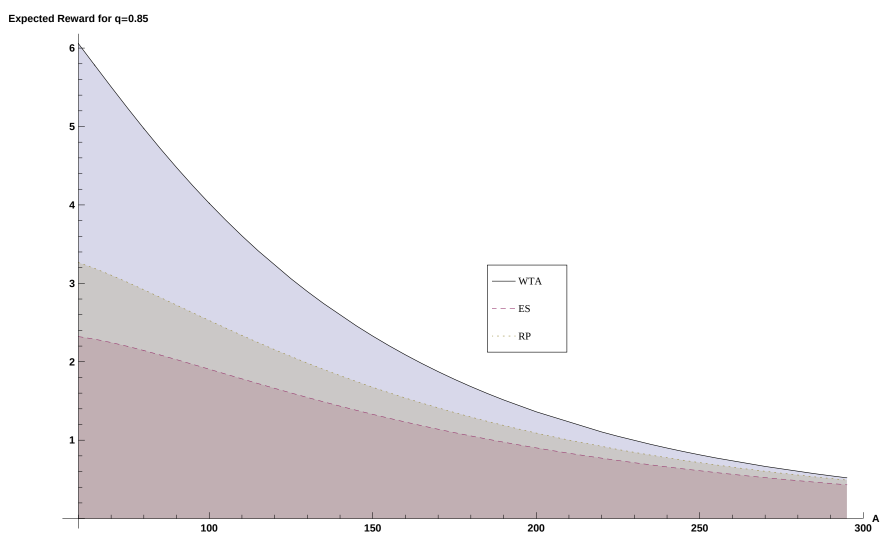


Figure 3: Total Reward (A) vs. Probability of Winning a Reward ($q = 0.6, n = 10,\ f = 1$).

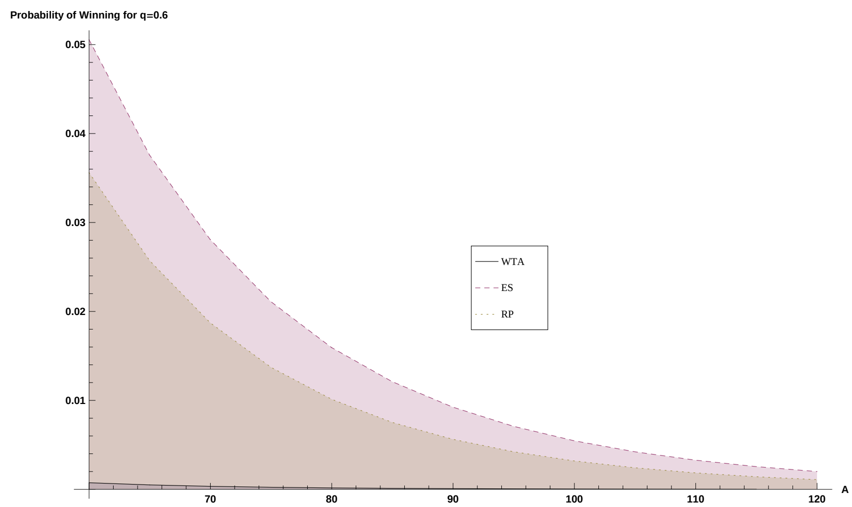


Figure 4: Total Reward (A) vs. Quality ($\frac{1+q}{2}$) ($n = 10,\ f = 1$)

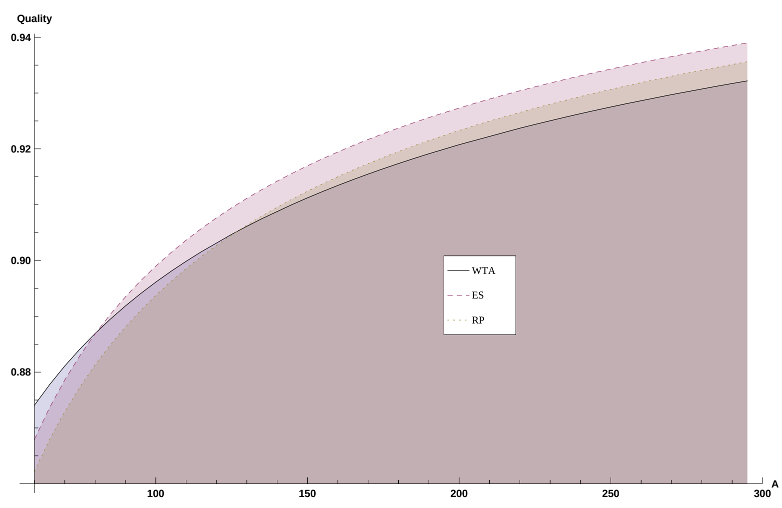


Figure 5: Total Reward (A) vs. Participation Rate ($1 - q$) ($n = 10,\ f = 1$).

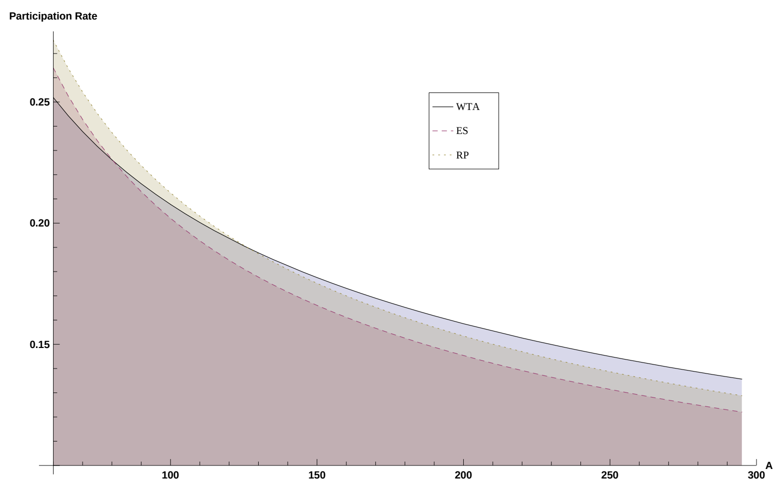


Figure 6: Total Reward (A) vs. Preferred Incentive ($n = 10,\ f = 1$)

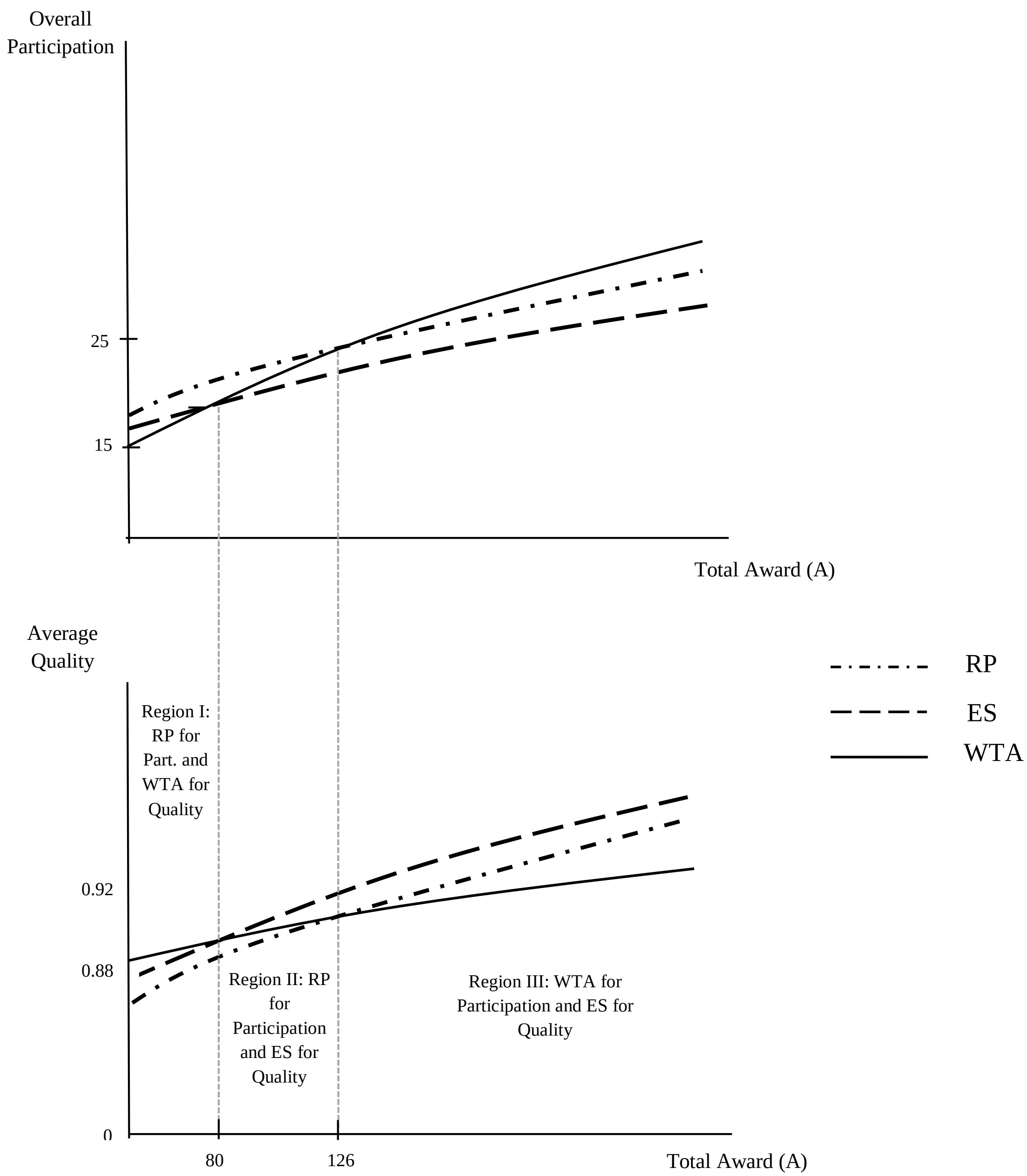


Figure 7: Histogram of Ratings in Contests with a Low Number of Entrants (left) and a High Number of Entrants (right)

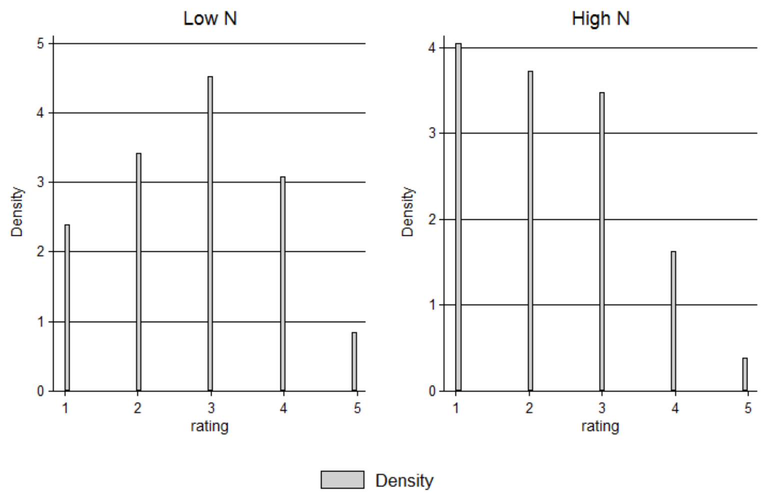


Figure 8: Out of Sample Predictions for Number of Contest Participants over Total Award

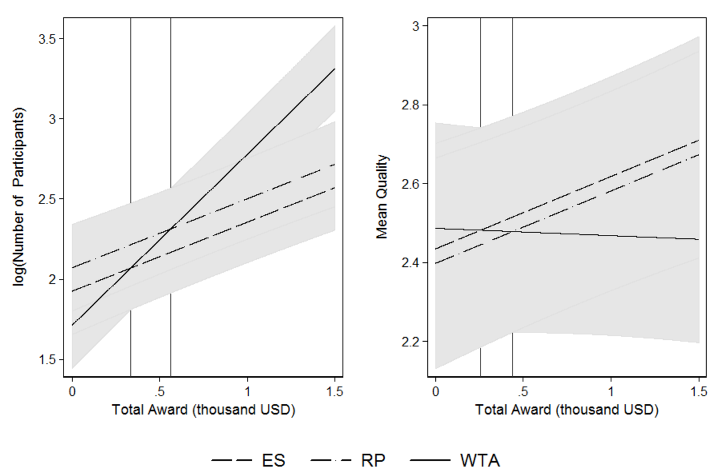


Table 1: Comparison of Incentives and Total Reward (A, for $f = 1$)

Seeker Objective	**Total Reward (A)**		
	Small	Medium	Large
Avg. Quality of Ideas	Winner Takes All (WTA)	Equally Shared (ES)	Equally Shared Reward (ES)
Overall Participation	Rank Proportional (RP)	Rank Proportional (RP)	Winner Takes All (WTA)

Table 2: Number of Contests and Reward Types by Contest-Category

Category	Subcategory	WTA	ES	RP	Total	%
Design-Print	Logo	9763	139	184	10086	62.89
Design-Print	Logo AND Stationery	1801	16	19	1836	11.45
Design-Print	Print design	969	35	36	1040	6.48
Design-Web	Small Website (uncoded)	833	7	13	853	5.32
Design-Print	Illustration	367	20	15	402	2.51
Writing-Creative	Company Naming	307	3	7	317	1.98
Design-Print	Clothing	181	28	18	227	1.42
Design-Web	Ad banner	177	30	16	223	1.39
Design-Print	Stationery	159	3	3	165	1.03
Design-Web	Landing Page (uncoded)	116	3	3	122	0.76
Design-Print	Package Graphics	108	5	4	117	0.73
Design-Web	Icons and Buttons	102	4	3	109	0.68
Design-Web	Header	73	3	1	77	0.48
Design-Web	Email Template (uncoded)	48	21	6	75	0.47
Design-Print	Presentation	59	0	1	60	0.37
Writing-Creative	Product Naming	49	2	0	51	0.32
Design-Web	Large Website (uncoded)	49	0	1	50	0.31
Writing-Business	Tagline	33	1	1	35	0.22
Design-Web	Widget and Apps (uncoded)	32	0	1	33	0.21
Writing-Business	Marketing Copy	24	0	0	24	0.15
Writing-Business	Article, Report & Proposal	22	0	0	22	0.14
Writing-Online	Web Content	13	1	5	19	0.12
Design-Mobile	Mobile App	15	0	0	15	0.09
Writing-Creative	Domain Naming	13	0	0	13	0.08
Design-Web	Blog Theme (uncoded)	12	0	0	12	0.07
Writing-Online	Newsletter	10	1	1	12	0.07
Design-Mobile	Mobile Icons	7	1	1	9	0.06
Design-Product	Packaging Design	5	0	0	5	0.03
Design-Product	Product Design	4	1	0	5	0.03
Writing-Creative	Essay & Short Story	3	0	1	4	0.02
Writing-Creative	Script Writing	4	0	0	4	0.02
Writing-Online	Online Marketing	4	0	0	4	0.02
Writing-Business	Resume	3	0	0	3	0.02
Design-Mobile	Mobile Illustration	1	1	0	2	0.01
Writing-Business	Editing and Proofreading	2	0	0	2	0.01
Writing-Online	Blog	2	0	0	2	0.01
Writing-Business	Book (business)	1	0	0	1	0.01
Writing-Business	Technical Writing	1	0	0	1	0.01
Writing-Creative	Book (creative)	1	0	0	1	0.01
Total		15373	325	340	16038	

Table 3: Mean Measures by Contest Design

	n	N	A (USD)	contests	submissions	submissions/solver
ES	2.526	31.252	826.90	325	114.182	4.339
RP	2.450	52.744	857.58	340	194.747	7.770
WTA	1	32.396	355.54	15,373	105.644	3.281
Mean/Total	1.062	32.804	375.73	16,038	107.706	3.398

Table 4: Hypotheses Testing

	Hypothesis	Test			
		coef	z	p $> \|z\|$	sig.
H_{1a}	for RP/ES: $\gamma_{13} > 0$	4.415	2.22	.026	**
H_{1a}	for WTA: $\gamma_{13} + \gamma_{14} > 0$	11.716	6.75	.000	***
H_{1b}	RP/ES: $\gamma_{23} > 0$	.203	3.81	.000	***
H_{1b}	WTA: $\gamma_{23} + \gamma_{24} > 0$	.041	1.10	.270	n.s.
H_{2a}	WTA > RP: $\gamma_{11} + \gamma_{14}A > 0$	18.219	2.09	.037	**
H_{2a}	WTA > ES: $\gamma_{11} + \gamma_{14}A - \gamma_{12} > 0$	26.967	3.19	.001	***
H_{2b}	RP > ES: $0 > \gamma_{12}$	-8.748	-3.91	.000	***
H_{2b}	RP > WTA: $0 > \gamma_{11} + \gamma_{14}A > 0$	-10.777	-4.75	.000	***
H_{2c}	RP > ES: $0 > \gamma_{12}$	-8.748	-3.91	.000	***
H_{3a}	ES > RP: $\gamma_{22} > 0$	.038	.62	.534	n.s.
H_{3a}	ES > WTA: $0 > \gamma_{21} + \gamma_{24}A - \gamma_{22} > 0$	-.550	-2.49	.013	**
H_{3b}	WTA > RP: $\gamma_{21} + \gamma_{24}A > 0$	.134	1.98	.047	**
H_{3b}	WTA > ES: $\gamma_{21} + \gamma_{24}A - \gamma_{22} > 0$	.096	1.48	.140	n.s.
H_{3c}	ES > RP: $\gamma_{22} > 0$	.038	.62	.534	n.s.

Table 5: Separate Estimation - Participation Equation

DV: $\dot{N}_{ij}$	Model 1		Model 2		Model 3	
	OLS		OLS		OLS	
A	6.929∗∗∗	(1.465)	5.473∗∗∗	(1.923)	4.415∗∗	(1.984)
WTA	−0.628	(.451)	−10.476∗∗∗	(2.222)	−10.777∗∗∗	(2.268)
WTA × A	−5.695∗∗∗	(2.071)	6.622∗∗∗	(2.476)	7.301∗∗∗	(2.483)
ES	−7.977∗∗∗	(1.827)	−8.191∗∗∗	(2.319)	−8.748∗∗∗	(2.239)
contestcount					.155∗∗∗	(.025)
prc_rated					−2.400∗∗∗	(.529)
featured					1.969∗∗	(.79)
duration					.226∗∗∗	(.043)
assured					2.733∗∗∗	(.844)
Category Controls	*no*		*yes*		*yes*	
Year Controls	*no*		*yes*		*yes*	
N	6563		6563		6563	
R^2	0.007		0.122		0.137	
LL	$-2.74E{+}04$		$-2.70E{+}04$		$-2.69E{+}04$	
AIC	54815.244		54031.189		53926.523	
BIC	54842.401		54133.027		54062.307	

SEs are clustered at the seeker level and in parentheses,* p<0.10, ** p<0.05, ***p<0.010.

Table 6: Separate Estimation - Quality equation

	Model 1 OLS		Model 2 OLS		Model 3 OLS		Model 4 HECKIT	
DV: $\dot{q}_{ij}^{avg}$	stdrate		stdrate		stdrate		stdrate	
A	0.029	(.046)	.156**	(.063)	.210***	(.058)	.203***	(.049)
WTA	−.060***	(.011)	.231***	(.064)	.148**	(.06)	.134**	(.06)
WTA × A	0.041	(.055)	−.214***	(.069)	−.175***	(.066)	−.163***	(.06)
ES	0.026	(.065)	0.106	(.071)	0.043	(.064)	0.038	(.057)
$\dot{N}_{ij}$					−.007***	(.001)	−.007***	(.001)
prcrated					−.161***	(.024)	−.138***	(.029)
contestcount					−0.001	(.001)	0	(.001)
featured					−.050**	(.024)	−.052**	(.026)
duration					0	(.001)	0	(.002)
assured					−0.032	(.027)	−0.03	(.029)
Category Controls	*no*		yes		yes		yes	
Year Controls	*no*		yes		yes		yes	
selection model (probit)								
A							0.014	(.094)
WTA							−0.042	(.127)
WTA × A							0.122	(.115)
ES							0.1	(.13)
$\dot{N}_{ij}$							.006***	(.001)
featured							−.087**	(.037)
duration							−.004**	(.002)
assured							−0.031	(.046)
prc_rated							.738***	(.036)
contcnt							.408***	(.036)
constant							−.625*	(.336)
error covariance λ							.056**	−0.026
N	5696		5696		5696		16038	
R^2	0.005		0.027		0.077		*n.a.*	
LL	−4587.442		−4522.716		−4374.217		*n.a.*	
AIC	9182.883		9075.433		8790.435		*n.a.*	
BIC	9209.473		9175.146		8930.033		*n.a.*	

SEs are clustered (bootstrapped for model 4) at the seeker level and in parentheses,* p<0.10, ** p<0.05, ***p<0.010.

Table 7: Instrumental Variable and Simultaneous equation

	Model 1		Model 2	
	2SLS		FIML	
DV: $\dot{N}_{ij}$				
A	3.925∗	(2.18)	3.925∗	(2.007)
WTA	−11.779∗∗∗	(2.822)	−11.779∗∗∗	(2.547)
WTA × A	6.007∗∗	(2.604)	6.007∗∗	(2.523)
ES	−10.092∗∗∗	(2.487)	−10.092∗∗∗	(2.277)
contcnt	.119∗∗∗	(.024)	.119∗∗∗	(.014)
featured	1.292∗	(.769)	1.292	(.861)
duration	.154∗∗∗	(.043)	.154∗∗∗	(.048)
assured	3.013∗∗∗	(.929)	3.013∗∗∗	(.841)
prcrated	−2.970∗∗∗	(.77)	−2.970∗∗∗	(.831)
$lag.\dot{N}_{ij}$	−.463∗∗∗	(.038)	−.463∗∗∗	(.035)
DV: $\dot{q}_{ij}^{avg}$				
$\dot{N}_{ij}$	−.007∗∗∗	(.002)	−.007∗∗∗	(.001)
A	.197∗∗∗	(.073)	.197∗∗∗	(.066)
WTA	.135∗	(.076)	.135∗	(.074)
WTA × A	−.159∗∗	(.078)	−.158∗∗	(.075)
ES	0.056	(.079)	0.056	(.071)
contcnt	0	(.001)	0	(.001)
featured	−.051∗	(.028)	−.051∗	(.028)
duration	0.001	(.002)	0.001	(.002)
assured	−0.032	(.033)	−0.032	(.033)
prcrated	−.120∗∗∗	(.032)	−.120∗∗∗	(.034)
Category Controls	*yes*		*yes*	
Year Controls	*yes*		*yes*	
error covariance $\sigma 12$			$-.020(n.s.)$	
N	3665		3665	
R^2	0.364		0.364	
LL	$-1.46E{+}04$		$-1.74E{+}04$	
AIC	29147.563		34968.67	
BIC	29277.901		35229.346	

Instrument: $lag.\dot{N}_{ij}$

Kleibergen-Paap rk LM statistic: 98.135

Kleibergen-Paap rk Wald F statistic: 188.078 (critical value 16.38)

SEs are clustered (bootstrapped for model 2) at the seeker level and in parentheses, * p<0.10, ** p<0.05, ***p<0.010.