

## Working Paper

---

# The Role of Feedback in Dynamic Crowdsourcing Contests: A Structural Empirical Analysis

Zhaohui (Zoey) Jiang

Stephen M. Ross School of Business  
University of Michigan

Yan Huang

Stephen M. Ross School of Business  
University of Michigan

Damian R. Beil

Stephen M. Ross School of Business  
University of Michigan

Ross School of Business Working Paper Series  
Working Paper No. 1334  
December 2016

This paper can be downloaded without charge from the  
Social Sciences Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=2884922>

# The Role of Feedback in Dynamic Crowdsourcing Contests: A Structural Empirical Analysis

Zhaohui (Zoey) Jiang, Yan Huang, Damian R. Beil

Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109

{jiangzh, yphuang, dbeil}@umich.edu

This Version: December 2016

In this paper, we empirically examine the impact of performance feedback on the outcome of crowdsourcing contests. We develop a dynamic structural model to capture the economic processes that drive contest participants' behavior, and estimate the model using a rich data set collected from a major online crowdsourcing design platform. The model captures key features of the crowdsourcing context, including a large participant pool, entries by new participants throughout the contest, exploitation (revision of previous submissions) and exploration (radically novel submissions) behaviors by contest incumbents, and the participants' strategic choice among these entry, exploration, and exploitation decisions in a dynamic game. We find that the cost associated with exploratory actions is higher than the cost associated with exploitative actions. High-performers prefer the exploitative strategy, while low-performers tend to make fewer follow-up submissions and prefer the exploratory strategy. Using counter-factual simulations, we compare the outcome of crowdsourcing contests under alternative feedback disclosure policies and award levels. Our simulation results suggest that the full feedback policy (providing feedback throughout the contest) may not be optimal. The late feedback policy (providing feedback only in the second half of the contest) leads to a better overall contest outcome.

*Key words:* Crowdsourcing contests, Feedback, Econometric analysis, Structural modeling, Dynamic game

---

## 1. Introduction

Recent technological advances have enabled organizations to better engage customers, freelance service providers, and independent experts in various business processes, such as customer information acquisition, product development, sourcing, and innovation. A wide variety of large-scale online platforms have emerged as platforms to facilitate these “crowdsourcing” applications. Depending on the design of such platforms, crowd workers or service providers may work independently, competitively, or cooperatively. *Crowdsourcing contests* are a popular form of competitive crowdsourcing; they often used to source innovations, and this use is the focus of our paper. Compared

to traditional innovation sourcing approaches, crowdsourcing contests allow an innovation seeker to access and interact with a much larger pool of innovators, to choose from a large number of submissions, and to pay only for the successful ones. This new way of sourcing innovation increases the variety and novelty of innovations, and lowers the risk of innovation failure. Recognizing the advantages of crowdsourcing contests, over the past several years, a rapidly increasing number of governments and corporations – including Google, Netflix, and even NASA – hosted crowdsourcing contests to leverage external human resources for sourcing innovation. There are also emerging crowdsourcing contest platforms, such as Innocentive, TopCoder, Kaggle, and 99Design, which serve as intermediaries between innovation seekers and crowd-innovators.

To be effective, crowdsourcing contests must be carefully designed and managed. The choices of award structure, competition size and problem specification can all influence the effectiveness of a crowdsourcing contest. Another important, but somewhat overlooked design element is the feedback disclosure policy. Many crowdsourcing contests are dynamic in nature – new solvers can join the contest at any time before the contest ends, and solvers who have made submissions can revise their existing submissions (exploit existing ideas) or submit new ideas/solutions (exploring new possibilities). The innovation seeker can in turn provide real-time feedback on the performance of existing submissions during a contest. Such performance feedback is an available feature on most crowdsourcing platforms. For example, Kaggle contests provide public leader boards displaying rankings based on the performance of submitted algorithms on a testing data set; crowdsourcing design platforms, 99design and Crowdspring, allow design-seekers to rate submitted designs using a 5-star rating system, and the rating distribution is publicly viewable. Most of these platforms strongly encourage seekers to provide feedback to contest participants: for example, the FAQ site of one platform states “the more feedback you can give, the better!” and “be sure to score EVERY entry.” Such recommendations are primarily based on the intuition that feedback can guide participants’ submissions towards seekers’ preference; as another platform states, “star ratings show how much the contest holder likes or dislikes a design”.

However, the effect of feedback is likely to be more complicated than that. In addition to conveying seekers’ preference, performance feedback also discloses the current status of the competition, and thus is likely to affect existing and potential contest participants’ participation decisions. Dating back to Fudenberg et al. (1983), theoretical literature on feedback in small-scale (mostly two-player) contests has generally concluded that the revealed performance gap will lead to both the low performer and the high performer reducing their efforts due to decreased competition. But this conclusion may not hold in crowdsourcing contests, because crowdsourcing contests are large-scale; new entrants can join the contest throughout the contest period; the notion of quality

can be quite subjective; existing creators can choose between exploitative and exploratory innovation strategies; and there are often multiple high performers competing against each other. In the presence of these unique features of crowdsourcing contests, revealing performance gaps may encourage participants to exert effort: new entrants are encouraged to join the contest when the feedback indicates a low level of competition; top performers want to secure their leading position; and low performers may endeavor to catch up – due to the subjectivity of the notion of “quality” in many crowdsourcing contests, innovators whose current submissions fall far behind have the option of submitting a completely new innovation, which gives them a chance to leapfrog the competition.

Given the co-existence of all these effects, there is no easy answer to the important question of whether (and if so when) to release performance feedback during crowdsourcing contests. The goal of this study is to examine this complicated process and disentangle the intertwining effects that in-process feedback may have on the outcome of crowdsourcing contests. More specifically, this study attempts to address the following research questions: (1) How does performance feedback affect existing contest participants’ follow-up innovation actions and potential entrants’ entry decisions? (2) What is the impact of the availability of feedback on the contest outcome, in terms of the total number of innovators participating in the contest, the highest quality achieved by their submissions, and the number of top performers/submissions? Does the timing of the feedback availability matter or not? (3) What is the value (disvalue) of providing feedback in crowdsourcing contests?

To answer these questions, we construct a dynamic structural model to analyze the effect of performance feedback on creators’ participation behavior in graphic design contests. The model explicitly captures how potential entrants decide whether to join an on-going contest, and how incumbents decide whether to make additional submissions, and if so, how to choose between the exploratory and exploitative strategies. We apply the structural model to a data set collected from a major platform for crowdsourced custom logo designs. In order to classify observed incumbents’ follow-up submissions into exploratory action (*redesign*) and exploitative action (*revision*), we employ the Scale-Invariant Feature Transform (SIFT) algorithm to quantify the similarity between each pair of submissions made by the same creator. We estimate the parameters governing creators’ (designers’) participation behavior, and find the following results. The cost associated with exploratory actions (*redesigns*) is higher than the cost associated with exploitative actions (*revisions*). High-performers prefer the exploitative strategy, while low-performers tend to make fewer follow-up submissions and prefer the exploratory strategy. In that sense, feedback not only helps encourage the leading creators, but also helps individual creators choose the most productive follow-up action and hone, or revamp, their designs according to how well the seeker likes, or dislikes, them.

Using the estimated parameters, we conduct counterfactual simulations to compare contest outcomes across four feedback policies: full/early/late/no feedback, where the seeker provides feedback throughout/only early/only late/not at all in the contest. The simulation results reveal that the maximum quality achieved is higher in the full feedback and the late feedback policies than in the no feedback and the early feedback policies. Between the full feedback and the late feedback policies, the difference in the maximum quality achieved is very small. Moreover, compared to no feedback, the same maximum submission quality can be achieved with  $2/3$  the award level, if the seeker provides feedback, even only in the second half of the contest. If the seeker’s objective is to maximize the number of top creators who achieve high ratings (5-star), or to maximize the total number of contest participants, the late feedback policy is in fact a better option than the full feedback policy. The full feedback policy performs worse in these two dimensions than the late feedback and the early feedback policies when the contest award is relatively high, and performs as poorly as the no feedback policy when the contest award is very high. These findings suggest that the common intuition, “the more frequently the seeker provides feedback the better”, may be misleading, even before we consider the cost the seeker incurs to evaluate submissions in real time to be able to provide feedback throughout the contest.

Our study makes several contributions. First, to our knowledge, this is the one of the first studies to investigate the impact of feedback on the outcome of crowdsourcing contests, and the first study that proposes a structural model to empirically analyze this impact. Second, the structural model presented in this paper is one of the most comprehensive models of crowdsourcing contests, as it explicitly models multiple stages, endogenous entry, and creators’ exploitation and exploration innovation strategies, unique features of crowdsourcing contests that, to our knowledge, have not been all incorporated in the existing literature. Third, our policy simulations yield very interesting results, which not only quantify (in dollar terms) the value of the feedback, but also provide rich managerial insights into whether, and if so, when should feedback be revealed, depending on the objective(s) that the seeker wants to achieve.

The rest of the paper proceeds as follows. Section 2 reviews relevant literature. Section 3 introduces the research context, describes the data used in the empirical analysis, and shows reduced-form evidence. We then present the dynamic structural model in Section 4. Section 5 discusses the estimation procedure and the estimation results. Section 6 describes how we conduct counterfactual experiments and the results from them. We summarize and conclude the paper in Section 7.

## **2. Literature Review**

Emerging large-scale online markets have attracted increasing academic interest from the operations management community. Researchers have studied the use of large-scale online markets

in different application contexts, such as acquiring customer preferences (Marinesi and Girotra 2013) and forecasting sales and commodity price (Bassamboo et al. 2015) using crowd-based tools, screening and managing human resources in large-scale online service marketplaces (Allon et al. 2012, 2014, Moreno and Terwiesch 2014), and raising funds from crowd donors (Hu et al. 2015, Hu and Wang 2015). On the other hand, there is also a large literature in operations management examining issues related to sourcing and procurement auctions (see Elmaghraby (2007) for a detailed review). Our paper studies another application of large-scale online markets, crowdsourced innovation or open innovation, which creates a new way of sourcing innovative products/ideas from a large group of creators/innovators (Girotra et al. 2010, Erat and Krishnan 2012). More specifically, our focus is the *crowdsourcing contest*, in which a large crowd of innovators compete to solve innovation-related problems.

Crowdsourcing contests have been examined as platforms for sourcing innovation (Terwiesch and Xu 2008, Terwiesch and Ulrich 2009). Existing literature on the design of crowdsourcing contests has looked at the impact of *award structure* (Moldovanu and Sela 2001, Yang et al. 2009, Liu et al. 2014); *optimal competition size* (Taylor 1995, Fullerton and McAfee 1999, Che and Gale 2003, Terwiesch and Xu 2008, Boudreau et al. 2011, Körpeoğlu and Cho 2015, Ales et al. 2016a, Boudreau et al. 2016); *problem specification* (Boudreau et al. 2011, Erat and Krishnan 2012); and *the disclosure of intermediate solutions* (Boudreau and Lakhani 2015, Wooten and Ulrich 2016b) on the outcome (both the quality and quantity of the crowdsourced solutions) of crowdsourcing contests. Most of the previous research studies crowdsourcing contests as static problems; however, in reality, a large portion of the crowdsourcing contests are dynamic in nature. Our paper explicitly models the dynamics of crowdsourcing contests, and focuses specifically on an important but rarely explored design element – feedback. To our knowledge, the only other study that looks at the role of feedback in crowdsourcing contest is Wooten and Ulrich (2016a), in which the authors conduct a field experiment to compare the performance of contests under three different feedback treatments – no feedback, random feedback, and truthful feedback. However, this paper does not explicitly model the contest dynamics, study the detailed mechanisms driving participants’ behavior, or examine the impact of feedback timing on contest outcomes.

Feedback in a different but related context, tournaments/contests, has been studied in the economics literature. The theoretical research focuses on small-scale contests, and until recently, most studies in this stream of research conclude that performance feedback reduces contestants’ effort, because when one player has a lead over the others, the subsequent contest will be biased in favor of the leading contestant and the followers cannot catch up with the leader by making the same level of effort (Fudenberg et al. 1983, Schotter and Weigelt 1992). A few recent papers (e.g., Gershkov and Perry 2009, Aoyagi 2010, Ederer 2010, Goltsman and Mukherjee 2011) extend the literature

by showing that the optimal feedback mechanism depends on the curvature of the agents’ cost function, complementarity between effort and ability, complementarity between contestants’ effort, etc. There are also laboratory experimental studies evaluating the role of feedback in contests. These studies have mixed results in terms of whether and how feedback will affect top-performing and low-performing contestants’ effort provision in various experimental settings (see Dechenaux et al. (2015) for a detailed summary of this literature). There has been very little work featuring feedback in contests for innovation. To our knowledge, there are only a few recent game-theoretic papers on this topic. Bimpikis et al. (2014) and Halac et al. (2016) both study “innovation races” and consider technical uncertainties in innovation races – namely whether it is feasible to solve the problem at all. They illustrate that feedback on the one hand exposes a discouraging performance gap, but on the other hand updates contestants’ perceptions of the feasibility of solving the problem. Mihm and Schlapp (2015) evaluates the average design quality and the best design quality under no feedback, public feedback and private feedback scenarios. They find that the best feedback strategy depends on the contest uncertainty and the contest holder’s interest in average design quality or the best design quality. A recent working paper by Gross (2016a) is one of the few papers that use field data to examine the effect of feedback on small-scale, fixed-size innovation contests. Concurrently but independently from our work, using a crowdsourcing data set similar to ours, the author estimates how feedback affects contestants’ participation and the quality of their subsequent submissions, in a model that treats each submission by the same participant as an independent trial. Then the author uses the estimated model to simulate the dynamics of a three-player, sequential-play contest, under alternative feedback policies, including the public, private and partial feedback policies. The simulation results suggest that the net effect of feedback on the number of high-quality submissions is positive; therefore, the author concludes that feedback is desirable for a principal seeking innovation.

Our work considers a similar principal decision (feedback), but in a different setting – crowdsourcing contests held on large-scale online markets. The crowdsourcing contests we study differ from small-scale fixed-size contests in the following crucial aspects: (1) the size of the participant pool is large; (2) crowdsourcing contests allow endogenous entries of new participants throughout the contest; and (3) there is large uncertainty in the relationship between effort and solution quality, which depends highly on the contest holder’s preferences. The last feature gives rise to exploration- and exploitation-type strategies; indeed, sourcing the best solution to an innovation problem has been modeled as a search process (Dahan and Mendelson 2001, Terwiesch and Loch 2004, Girotra et al. 2010, Kornish and Ulrich 2011, Erat and Krishnan 2012, Ales et al. 2016b, Gross 2016b), where independent trials (exploration) and sequential trials (exploitation) contribute differently to the contest outcomes. In the presence of these unique features, we expect that the role of feedback

will be quite different in crowdsourcing contests, and that findings in the literature about the role of feedback in traditional contests may not hold in the context of crowdsourcing contests.

The model we propose in this paper captures these unique features of crowdsourcing contests by explicitly modeling potential entrants’ entry decisions and incumbents’ choices between exploratory and exploitative strategies in this highly uncertain environment, and, as a result, provides novel insights into the role of feedback in crowdsourcing contests. In addition, our study contributes to the limited empirical literature on the design of crowdsourcing contests and on the sourcing of innovative products. Using a structural modeling approach, we are able to recover the parameters governing contest participants’ behavior from real-world data and conduct counterfactual simulations to evaluate alternative feedback disclosure structures. Although the focus of our study is the role of feedback, our structural model of contest participants’ behavior can be used to study other features of crowdsourcing contests and other large-scale online open platforms.

Methodologically, our paper is based on the dynamic game structural estimation literature (see Aguirregabiria and Mira (2010) for a detailed review). Our modeling and estimation approach diverges from the more conventional framework where the market is assumed to be in a stationary environment and the competition has an infinite horizon. Specifically, we embed discrete choice with private information into a non-stationary, finite-horizon dynamic game, focus on type-symmetric strategies to avoid multiple equilibria, and use Rust (1987)’s nested fixed-point (NFXP) estimation approach to recover the parameters that govern creators’ decision making processes. Our work contributes to the growing empirical operations management literature that employs structural modeling approach to examine operations-related questions, such as pricing strategy (Li et al. 2014, Moon et al. 2016), service provision (Allon et al. 2011, Lu et al. 2013, Aksin et al. 2013, Guajardo et al. 2015, Yu et al. 2016, Xu et al. 2016), operational costs (Olivares et al. 2008, Musalem et al. 2010, Mani et al. 2015), and bullwhip effect in supply chains (Bray and Mendelson 2012, 2015). Since the crowdsourcing contests can be viewed as a form of all-pay auctions, this study is most closely related to Olivares et al. (2012), Kim et al. (2014), and Hyndman and Parmeter (2015), which use the structural modeling approach to study bidder behavior in various types of auctions.

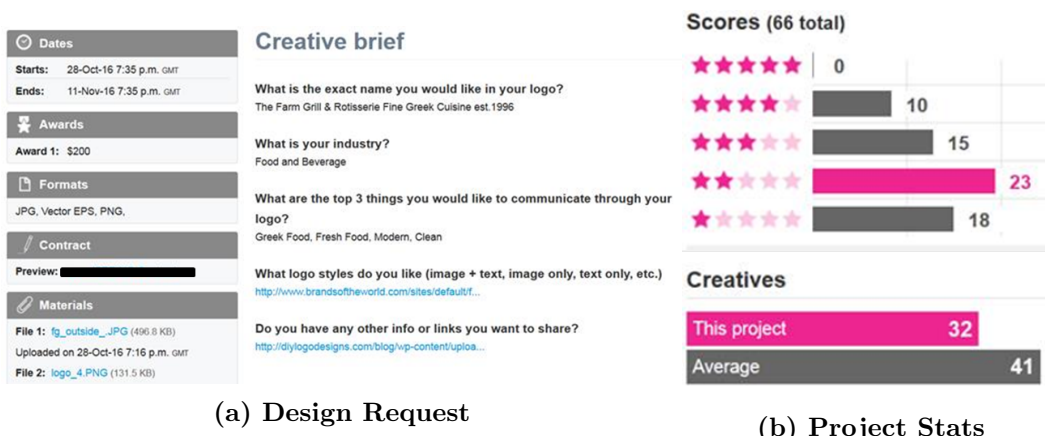
### **3. Research Context and Data**

Our data are collected from a major online platform for crowdsourcing creative services in various areas, including custom logo design, Web design, industrial design and writing services. We focus on logo design contests because it is the largest category both in terms of the number of completed contests and the number of active creators.

A typical logo design contest proceeds as follows. First, a seeker (“he”) in need of a design posts a design request. In the posting, he describes what he needs, specifies when he needs it (i.e., the



contest’s start date and end date), chooses whether he wants to make all existing submissions public by choosing the “public-gallery” option, and announces the award structure (i.e., whether the award is “assured<sup>1</sup>”, the number of winners, and the award(s) for the winner(s)). Once a contest is posted, all creators (“she”) on the platform can join the contest and submit design(s) at any time before the contest ends. At any time during the contest period, the seeker can rate each submission based on a 5-star system. A submission’s rating is visible to its author. All ratings are summarized in a table, called “Project Stats” (Figure 1), which is accessible to all participating and potential creators.<sup>2</sup> At the end of the contest, the seeker picks his favorite submission(s) and gives the pre-announced award to its (their) author(s).



**Figure 1 Illustrative Screenshots of the Design Request and Project Stats**

Note: “Buyer Assured”, “public-gallery”, “Award”, “Start-date and End-date” and “Project length” are labeled on the contest list page and on a project’s header.

We collect data of “public-gallery” logo-design contests on this crowdsourcing platform from March, 2012 to November, 2014. For each contest, we record all participating creators’ and the seeker’s activities (including creators’ submissions and the seeker’s ratings) with corresponding time stamps, and download all design images. To facilitate the empirical analysis, we focus on 7-day contests where the design seekers promise to reward one and only one final winner, because it has been documented (Moldovanu and Sela 2001, Yang et al. 2009, Liu et al. 2014) that the contest length and award structure can affect creators’ behavior and contest outcomes; since the objective of this study is to examine the role of feedback, we purposefully minimize the heterogeneity among the contests in these other dimensions. The contests included in our sample are representative

<sup>1</sup> In an “assured” contest the seeker promises to give out the pre-specified award.

<sup>2</sup> It is worth pointing out that creators only observe the distribution of the existing ratings, i.e., how many 1-star, 2-star, ..., 5-star submissions there are (e.g., there are eight 5-star submissions, forty 4-star submission in the example displayed in Figure 1), but do not know which submissions are rated as 5-star, which ones receive a 4-star rating, etc. This prevents creators from copying good designs.

contests on the platform – 97% of the contests held on the platform have a single award, 61% are “Buyer Assured”, and 7-day is the most common length among all contests. The final working sample consists of 810 contests, 26,367 contest-creator combinations, 75,572 design images and 45,999 ratings. Table 1 reports summary statistics of contest-level characteristics.

**Table 1 Contest-Level Summary Statistics**

Variables	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
<i>Award</i> (\$)	260.84	97.57	200	200	200	300	1,000
<i>No. Submissions</i>	93.30	73.39	14	56	77	109	1,221
<i>No. Creators</i>	32.55	19.81	6	20	29	39	233
<i>No. Submissions with Ratings</i>	56.79	63.13	0	19	43	73	927
<i>No. 1-Star Submissions</i>	7.14	21.78	0	0	1	6	411
<i>No. 2-Star Submissions</i>	10.82	17.03	0	1	6	14	240
<i>No. 3-Star Submissions</i>	17.10	22.85	0	4	11	23	347
<i>No. 4-Star Submissions</i>	11.00	13.37	0	3	7	15	141
<i>No. 5-Star Submissions</i>	3.16	5.74	0	0	1	4	48

### 3.1. Classifying Creators’ Actions

When exploring creators’ submission patterns in our data, we find that an existing creator’s follow-up submissions are sometimes similar to her previous submissions, while other times they are very different. The former type of follow-up submissions can be considered *revisions* of a creator’s previous submissions, akin to an “exploitative” innovation strategy; by contrast the latter type involves creating new design(s) that are significantly different from any of the creator’s existing designs (later defined as *redesigns*), corresponding to an “exploratory” innovation strategy.<sup>3</sup>

It is important to distinguish between these two types of submissions, as they have different implications from both the seeker’s perspective and the creators’ perspective. For the seeker, these different submission types contribute differently to the pool of designs seekers can choose from – a *revision* may result in an incremental quality difference relative to prior designs, whereas a radically different *redesign* may impart greater quality differences relative to previous designs and increases the variety of the submission portfolio. From a creator’s stand point, it could cost them different amounts of effort to make these two types of submissions: tweaking an existing design for a revision is likely to require less effort, while creating a radically different *redesign* will possibly require much more effort.

<sup>3</sup> There are also a small number of cases where a follow-up submission is almost identical to one of the creator’s previous submissions. We classify these submissions as *replications*, and in the later analysis, *replications* are not counted as follow-up submissions and are removed from the data, since we assume “replicating” an existing design incurs very little cost, and does not benefit the design seeker very much.

To systematically classify creators’ follow-up submissions observed in the data into exploitative actions (*revisions*) and exploratory actions (*redesigns*) on a large scale, we employ an image classification algorithm called *Scale-Invariant Feature Transform (SIFT)*. SIFT is an algorithm used to detect and describe local features in images proposed by Lowe (1999). The algorithm consists of four steps. First, from a pair of design images (A and B), we extract descriptors of the key points by identifying *SIFT feature vectors* in scale space, which robustly capture the structural properties of the images. As a second step, we match SIFT feature vectors by calculating and comparing the Euclidean distance between each of the SIFT feature vectors in image A and image B. Using the obtained matched feature-vector pairs, we then calculate the Similarity Ratio (the percentage of matched SIFT features relative to the total number of SIFT features in images A and B). Finally, we classify the image pair (image A and image B) as either similar or different based on the Similarity Ratio. The higher the ratio is, the more similar the two images are. In our empirical analysis, we classify a pair of submissions as similar if the Similarity Ratio is greater or equal to 0.4. Correspondingly, if a creator’s new submission is very similar to any of her prior submissions (the similarity score between the two submissions is above 0.4), we classify the new submission a “revision”; otherwise, we consider the new submission as a “redesign”. (See Appendix A for a more detailed description of our algorithms and its performance.) To facilitate our empirical analysis, we will divide the contest time horizon in the data into discrete intervals, or *periods*. Then based on a creator’s submissions in a period being only “redesigns”, only “revisions”, or both “revisions” and “redesigns”, we define her follow-up action decision within that period as *redesign*, *revise*, or *do-both(revise-and-redesign)*, respectively. If she submitted nothing during the period, her action is *do-nothing*.

### 3.2. Reduced-Form Evidence

Before constructing our main structural model, we use regression analysis to explore how disclosed ratings are associated with creators’ participation behavior. The purpose is to gain preliminary insight from the data that we can then build upon with a more sophisticated and powerful structural analysis. Specifically, we divide creators into “new entrants” and “existing creators” / “incumbents”, and test separately whether the number of new entrants or incumbents’ follow-up actions are affected by the ratings that the seeker has disclosed by the end of the previous day (Equations 1 and 2, respectively). Incumbents’ follow-up actions in a period (day) are classified into *redesign*, *revise*, *do-both*, or *do-nothing* based on SIFT image comparison results from Section 3.1.

In the first regression (Equation 1), we regress the number of entrants joining contest  $q$  on day  $t$  ( $\Delta(\text{No. Creators})_{qt}$ ) on the numbers of 1-star, 2-star, ..., 5-star ratings disclosed up to day  $t - 1$  ( $(\text{No. 1-Star})_{qt-1}$ ,  $(\text{No. 2-Star})_{qt-1}, \dots$ ,  $(\text{No. 5-Star})_{qt-1}$ ), while controlling for the number of

creators already in the contest ( $(No. Creators)_{qt-1}$ ) and the cumulative number of submissions made by all existing participants ( $(No. Submissions)_{qt-1}$ ) up to day  $t - 1$ , as well as the time dummies and contest-level fixed effects.

In the second regression (Equation 2), we apply a multinomial logit regression model to incumbents' follow-up action choices. The dependent variable is a nominal variable denoting incumbent  $i$ 's choice  $Action_{iqt}$  among *redesign*, *revise*, *do-both*, or *do-nothing*, where the reference category is *do-nothing*. We include three main sets of independent variables in this multinomial logit regression: (1) the individual-level variables, including the number of submissions that the *focal* creator has made previously ( $(No. Submissions)_{iqt-1}$ ), and among all her previous submissions her best rating ( $(Best Rating)_{iqt-1}$ ), second-best rating ( $(SecondBest Rating)_{iqt-1}$ ), and average rating ( $(Avg Rating)_{iqt-1}$ ), (2) the contest-level rating variables, including  $(No. 1-Star)_{qt-1}$ ,  $(No. 2-Star)_{qt-1}$ , ...,  $(No. 5-Star)_{qt-1}$ , and (3) control variables, including  $(No. Submissions)_{qt-1}$  and  $(No. Creators)_{qt-1}$ . Additional controls include the amount of award for the contest ( $Award_q$  in \$) and time dummies. To simplify our notation, we group independent variables into three vectors

$$\begin{aligned}\mathbf{W}_{qt} &:= \{(No. Creators)_{qt}, (No. Submissions)_{qt}\}; \\ \mathbf{Y}_{qt} &:= \{1, (No. 1-Star)_{qt}, (No. 2-Star)_{qt}, \dots, (No. 5-Star)_{qt}\}; \\ \mathbf{Z}_{iqt} &:= \{(No. Submissions)_{iqt}, (Avg Rating)_{iqt}, (Best Rating)_{iqt}, (SecondBest Rating)_{iqt}\}.\end{aligned}$$

$$\Delta(No. Creators)_{qt} = \beta \mathbf{Y}_{qt-1} + \Psi \mathbf{W}_{qt-1} + \phi_q + \delta_t + \mu_{qt} \quad (1)$$

$$\ln \frac{Pr(Action_{iqt} = k)}{Pr(Action_{iqt} = \text{do-nothing})} = \Gamma_k \mathbf{Z}_{iqt-1} + \Lambda_k \mathbf{Y}_{qt-1} + \alpha_k \mathbf{W}_{qt-1} + \zeta_k Award_q + \rho_t + \nu_{iqt}, \quad (2)$$

where  $k = \text{redesign, revise, or do-both}$

The estimated coefficients reported in Table 2 suggest that, after controlling for  $(No. Creators)_{qt-1}$ ,  $(No. Submissions)_{qt-1}$ , and the contest-level fixed effects, a larger number of high ratings (5-star) disclosed in previous periods is associated with fewer entries, whereas a larger number of low ratings (1-star and 2-star) is associated with more entries. In addition, the number of existing participants is negatively correlated with the number of entries, which suggests that when a contest is already "crowded", potential entrants are discouraged from joining the contest.

The estimation results reported in Table 3 indicate that incumbents' own best rating is positively correlated with the probabilities of them choosing the *redesign*, *revise* and *do-both* actions, and this correlation is the highest with the *revise* action. Additionally, once we control for the focal incumbent's best rating, neither her average rating nor her second-best rating is significantly associated with any of the action probabilities. Further, when there are fewer creators or fewer

**Table 2 Regression of the Number of Entries**

Dependent Variable: $\Delta Creator_{qt}$		
$(No. 1-Star)_{qt-1}$	0.026**	(0.008)
$(No. 2-Star)_{qt-1}$	0.022*	(0.010)
$(No. 3-Star)_{qt-1}$	0.000	(0.008)
$(No. 4-Star)_{qt-1}$	-0.021	(0.012)
$(No. 5-Star)_{qt-1}$	-0.149***	(0.022)
Control Variables		
$(No. Submissions)_{qt-1}$	-0.004	(0.006)
$(No. Creators)_{qt-1}$	-0.202***	(0.015)
Individual-Level Fixed Effects	Yes	
Time Dummies	Yes	
Observations	5,607	
R <sup>2</sup>	0.166	
Adjusted R <sup>2</sup>	0.142	
F Statistic	136.369***	(df = 7)

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001; the numbers in parenthesis are standard errors

**Table 3 Multinomial Logit Regression of the Incumbent Follow-up Action Choice**

	Depend Variable: $Action_{iqt}$		
	<i>re-design</i>	<i>revision</i>	<i>do-both</i>
Individual-Level Variables			
$(No. Submissions)_{iqt-1}$	0.011 (0.014)	0.080*** (0.006)	0.075*** (0.009)
$AvgRating_{iqt-1}$	0.099 (0.135)	0.101 (0.074)	0.139 (0.120)
$BestRating_{iqt-1}$	0.269*** (0.081)	0.352*** (0.048)	0.188* (0.076)
$SecondBestRating_{iqt-1}$	-0.095 (0.088)	-0.033 (0.051)	-0.076 (0.081)
Contest-Level Variables			
$Award_q(\$)$	0.001 (0.000)	0.001*** (0.000)	0.001*** (0.000)
$(No. 1-Star)_{qt-1}$	0.010*** (0.003)	0.010*** (0.002)	0.011*** (0.003)
$(No. 2-Star)_{qt-1}$	0.009** (0.003)	0.008*** (0.002)	0.012*** (0.003)
$(No. 3-Star)_{qt-1}$	0.005 (0.003)	0.008*** (0.002)	0.006* (0.003)
$(No. 4-Star)_{qt-1}$	0.002 (0.004)	0.005* (0.003)	0.014*** (0.004)
$(No. 5-Star)_{qt-1}$	-0.012 (0.009)	-0.027*** (0.005)	-0.030*** (0.008)
$(No. Submissions)_{qt-1}$	-0.007* (0.003)	-0.002 (0.002)	-0.007* (0.003)
$(No. Creators)_{qt-1}$	0.005 (0.006)	-0.018*** (0.004)	-0.009 (0.006)
Time Dummies	Yes		
Observations	24,085	Log Likelihood	-14,783.050
R <sup>2</sup>	0.040	LR Test	1,247.014*** (df = 54)

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001; the numbers in parenthesis are standard errors.

existing submissions in the contest, and when there are more low ratings and fewer high ratings disclosed, the focal player is more likely to follow-up with a non-null action.

The reduced-form analysis discussed above provides evidence for the correlation between performance feedback and the participation behavior of both potential entrants and existing participants. However, these regression models have not fully captured the dynamic nature of the contest and the interaction among creators. Here we have only detected the potential effect of performance feedback

on the individual participation behavior of either a potential entrant or an incumbent, but have not been able to test, or imply, the relationship between feedback and overall contest outcomes. For example, the reduced-form results suggest that revealing high ratings can encourage the authors of those highly-rated designs to improve upon their own submissions; but on the other hand, it discourages other participants' submission activities and entries. We cannot directly compare the magnitudes in the regression coefficients to see which one dominates another. In addition, in the data, seekers in most contests follow the platform's suggestion and provide feedback throughout the entire contest, and as a result, there is not enough variation in the feedback disclosure policy that would allow us to directly compare the outcome of contests with different feedback disclosure policies. Even if we had such variations in the data, policy inferences from the reduced-form results could be misleading because any policy change that constitutes a major regime shift that alters the key elements of the decision process (which is likely to be the case in our context, as changes in the feedback disclosure policy will alter participants' information and belief structure) can potentially lead to unstable responses (Lucas 1976). By delving into the underlying decision-making mechanism and explicitly modeling the decision primitives, we will be able to more reliably evaluate alternative feedback disclosure policies, and measure the effect of performance feedback on contest dynamics and outcomes. Therefore, we develop a structural model of creators' participation behavior, which is explained in detail in the next section.

#### 4. The Structural Model

Based on the reduced-form evidence from Section 3, we build a finite-horizon dynamic game model to capture creators' behavioral dynamics in a crowdsourcing design contest. Time is divided into periods and indexed by  $t$ . In period  $t$ , the set of potential entrants who may choose to join the contest and the set of incumbent creators are denoted as  $M_t$  and  $N_t$  respectively.<sup>4</sup> We index potential entrants by  $j$ , and incumbent creators by  $i$ . We also attach superscript  $e$  to potential entrants' response functions to distinguish them from incumbents' functions. In the main model, the seeker commits to the full feedback policy (he will provide feedback throughout the contest), as most seekers (>70%) provide full feedback in our sample. In our model, the seeker provides his feedback at the end of each period; this reflects the reality that the seeker is not immediately able to give feedback on a one-for-one basis as each individual submission rolls in. The timing of the model is as follows:

1. The contest begins with the seeker posting a design request and announcing the award ( $R$ ).
2. In the first period ( $t = 1$ ), potential entrants ( $M_1$ ) arrive at the contest, and decide whether to join the contest by submitting their first design(s). Potential entrants' decision in each

<sup>4</sup> We assume the pool of potential entrants is renewed every period.

period (including the first period and periods  $2, \dots, T$ ) is a binary choice, denoted as  $d_{j,t} \in \{\text{enter}, \text{not-enter}\} := \mathcal{D}$ . The seeker evaluates the submitted designs, and at the end of the period, discloses the ratings. The rating is on the scale of 1-5 stars; if a design is not rated, we use “NA” to denote its rating.

3. At the beginning of each of the subsequent periods ( $t = 2, \dots, T$ ), incumbents ( $N_t$ ) and potential entrants ( $M_t$ ) observe all existing ratings. Based on this information, they make the following decisions simultaneously: potential entrants decide whether to join the contest ( $d_{j,t}$  defined above), and incumbents choose their follow-up action  $a_{i,t} \in \{\text{do-nothing}, \text{redesign}, \text{revise}, \text{do-both}\} := \mathcal{A}$ , where,  $a_{i,t} = \text{do-nothing}$ , if incumbent  $i$  decides not to do anything in period  $t$ ;  $a_{i,t} = \text{redesign}$ , if she creates one or more new designs that are significantly different from any of her existing designs;  $a_{i,t} = \text{revise}$ , if she revises one or more of her existing designs;  $a_{i,t} = \text{do-both}$ , if she both revises one or more of her own existing designs and submits one or more brand new designs.
4. In period  $T + 1$ , the creator of the best quality design wins the contest and receives the prize ( $R$ ). In this terminal period, no entry is allowed, and incumbents do not have a chance to take any action.<sup>5</sup>

Since the competition is quality-based and ratings reflect submissions’ quality, existing ratings can be considered as state variables in our model, which not only capture the current status of the contest, affect creators’ current and future utility, but also evolve as a function of creators’ actions in each period. However, in reality, it is difficult for creators to track the ratings of each one of their own submissions and their competitors’ submissions; as we can see from the reduced-form results,  $BestRating_{iqt-1}$  significantly affects creator  $i$ ’s follow-up actions, while  $SecondBestRating_{iqt-1}$  and  $AvgRating_{iqt-1}$  do not (Table 3). Moreover, incorporating all submissions’ ratings and their evolution will make the model unmanageable. Hence, we define the individual-level state variable at time  $t$  (denoted as  $x_{i,t}$ ) as the highest rating received by an incumbent  $i$  up to the beginning of period  $t$ .<sup>6</sup> Correspondingly, we define the vector  $s_t = \{s_t(x)\}_{x \in \{\text{NA}, 1, 2, 3, 4, 5\}}$  as the contest-level state variable, where  $s_t(x)$  is the number of incumbents whose individual state takes value  $x$  in period  $t$ .

The state variables  $x_{i,t}$  and  $s_t$  evolve as follows. A contest starts with zero incumbents and a contest state  $s_1 = \{0, 0, 0, 0, 0, 0\}$ . At the beginning of each period  $t$  where  $t \in \{1, 2, \dots, T\}$ , potential entrants arrive and make entry decisions. If a potential entrant  $j$  enters the contest in period  $t$ , she will become an incumbent from  $t + 1$ , and the best rating her submitted design(s) receives in

<sup>5</sup> We add an arbitrary period  $T + 1$  at the end of the model to represent the time when the winner is announced.

<sup>6</sup> We use a creator’s best rating observed on a 5-star scale as her individual-level state variable for now. One potential problem with this definition is that the ratings are truncated at 5. If a creator already has a 5-star design but revises her design, her best rating cannot be improved any further. To deal with this problem, we slightly alter the definition of the individual-level state variable, which will be explained later in this section.

period  $t$  will become her individual state at the beginning of the next period ( $x_{j,t+1}$ ), which is a random draw from the probability distribution  $p^e(x_{j,t+1})$ . For any incumbent  $i$ , her state variable  $x_{i,t}$  evolves as a function of her action  $a_{i,t}$ . The action-specific transition probability of the state variable for an incumbent creator is then  $p(x_{i,t+1}|x_{i,t}, a_{i,t})$ . The contest-level state variable,  $s_t$ , will evolve correspondingly, which can be expressed as  $p(s_{t+1}|s_t, a_t, d_t)$ , where  $a_t$  is the stack of all incumbents' actions in period  $t$ , and  $d_t$  is the stack of all potential entrants' entry decisions in period  $t$ . Note that  $p^e(\cdot)$ ,  $p^e(\cdot|\cdot)$ , and  $s_t$  are common knowledge for all creators (Aguirregabiria and Mira 2010).

#### 4.1. Single Period Utility

As there are finitely many of periods in each contest, the per-period utility is  $t$ -dependent. We first explain the per-period utility in the terminal period  $T + 1$ . In period  $T + 1$ , no creator action is allowed; the creator with the best quality design is announced as the winner and receives award  $R$ ; everyone else receives nothing. Given creator  $i$ 's state  $x_{i,T+1}$  and the contest-level state  $s_{T+1}$  from the seeker ratings, creator  $i$ 's expected per-period utility in period  $T + 1$  can be expressed as

$$U_{i,T+1}(x_{i,T+1}, s_{T+1}) = \alpha R \cdot Pr(i \text{ wins} | x_{i,T+1}, s_{T+1}), \quad (3)$$

where  $\alpha$  is the marginal utility of money (or the number of utils a creator receives from getting an additional dollar of award). We will explain how to calculate  $Pr(i \text{ wins} | x_{i,T+1}, s_{T+1})$ , the probability that creator  $i$  wins a contest, later in this section.

In periods  $t = 1, \dots, T$ , incumbent  $i$ 's per-period utility can be expressed as

$$U_{i,t}(a_{i,t}, \epsilon_{i,t}) = -c(a_{i,t}) + \epsilon_{i,t}(a_{i,t}), \quad a_{i,t} \in \mathcal{A}, \quad (4)$$

where  $c(a_{i,t})$  represents the cost associated with action  $a_{i,t}$ , and  $\epsilon_{i,t}(a_{i,t})$  represents the individual-level choice-specific random shock. Likewise, potential entrant  $j$ 's per-period utility can be expressed as

$$U_{j,t}^e(d_{j,t}, \epsilon_{j,t}) = -c_t^e(d_{j,t}) + \epsilon_{j,t}(d_{j,t}), \quad d_{j,t} \in \mathcal{D}, \quad (5)$$

where the  $c_t^e(d_{j,t})$  term represents the cost associated with action  $d_{j,t}$  in period  $t$ .  $c_t^e(d_{it} = \text{enter})$  may include the costs of becoming aware of the contest, understanding the problem specifications, and coming up with the first design(s), and thus can be different across periods.<sup>7</sup> For normalization purposes,  $c(a_{i,t} = \text{do-nothing})$  and  $c_t^e(d_{j,t} = \text{not-enter})$  are assumed to be zero. The shocks  $\epsilon_{i,t}(a_{i,t})$  ( $\epsilon_{j,t}(d_{j,t})$ ) is private information observable to incumbent  $i$  (potential entrant  $j$ ) in period  $t$  before

<sup>7</sup> Notice that we assume the cost associated with incumbents' actions ( $c(a_{i,t})$ ) to be time invariant, but allow entry cost ( $c_t^e(d_{j,t})$ ) to vary across periods, due to the fact that the crowdsourcing platform displays contests nearing their end higher on the contest list webpage, and as a result the cost of discovering a contest may decrease over time.



she chooses which action to take, but are unobservable to other creators and the researchers. In addition, the shocks  $\epsilon_{i,t}(a_{i,t})$  and  $\varepsilon_{j,t}(d_{j,t})$  are assumed to be zero-mean, and *i.i.d.* with respect to individual, time and choice.  $\epsilon_{i,t} = \{\epsilon_{i,t}(a_{i,t})|a_{i,t} \in \mathcal{A}\}$  follows a distribution whose probability density function is  $p_\epsilon$ , and  $E_{i,t}$  is the corresponding state space;  $\varepsilon_{j,t} = \{\varepsilon_{j,t}(d_{j,t})|d_{j,t} \in \mathcal{D}\}$  follows a distribution whose probability density function is  $p_\varepsilon$ , and  $\mathcal{E}_{j,t}$  is the corresponding state space.

## 4.2. Creators' Decisions

In every period  $t$ , incumbents decide on follow-up actions, and potential entrants make entry decisions. Creators are forward-looking and take into account the implications of their decisions on future utilities and on the expected future reaction of competitors. Specifically, creators make these decisions to maximize expected discounted lifetime utility.

We assume creators play Markov strategies.<sup>8</sup> Formally, a Markov strategy for incumbent creator  $i$  in period  $t$  is a function  $\rho_{i,t} : X_{i,t} \times S_t \times E_{i,t} \rightarrow \mathcal{A}$ . Likewise, a Markov strategy for potential entrant  $j$  in periods  $t$  is a function  $\lambda_{j,t} : S_t \times \mathcal{E}_{j,t} \rightarrow \mathcal{D}$ . Let  $\sigma_t = \{\rho_t, \lambda_t\}$  summarize all existing incumbents' and potential entrants' strategies in period  $t$ , where  $\rho_t = \{\rho_{i,t}\}_{i \in N_t}$  and  $\lambda_t = \{\lambda_{j,t}\}_{j \in M_t}$ .  $\sigma = \{\sigma_t\}_{t=1, \dots, T}$  then summarizes all periods' strategies. Note that the strategies are time-varying because the contest has a finite horizon. Let  $V_{i,t}^\sigma(x_{i,t}, s_t, \epsilon_{i,t})$  represent the value function for an incumbent  $i$  given that the other creators behave according to their respective strategies in  $\sigma$ , and given that incumbent  $i$  uses her best response strategy. That is,

$$V_{i,t}^\sigma(x_{i,t}, s_t, \epsilon_{i,t}) = \max_{a_{i,t}} \mathbb{E} \left\{ \left[ \sum_{\tau=t}^T \beta^{\tau-t} U_{i,\tau}(a_{i,\tau}, \epsilon_{i,\tau}) + \beta^{T+1-t} U_{i,T+1}(x_{i,T+1}, s_{T+1}) \right] \middle| x_{i,t}, s_t, \epsilon_{i,t}; \sigma \right\}. \quad (6)$$

The expectation is taken over current-period private shocks of other contestants, as well as future values of the state variables and private shocks. According to Bellman's principle of optimality, we can write  $V_{i,t}^\sigma(x_{i,t}, s_t, \epsilon_{i,t})$  recursively as:

$$V_{i,t}^\sigma(x_{i,t}, s_t, \epsilon_{i,t}) = \begin{cases} \max_{a_{i,t}} \left\{ U_{i,t}(a_{i,t}, \epsilon_{i,t}) + \beta \mathbb{E}[V_{i,t+1}^\sigma(x_{i,t+1}, s_{t+1}, \epsilon_{i,t+1})|a_{i,t}] \right\} & \text{if } t \in 1, 2, \dots, T, \\ U_{i,T+1}(x_{i,T+1}, s_{T+1}) & \text{if } t = T + 1; \end{cases} \quad (7)$$

and for potential entrants (in  $t = 1, \dots, T$ ),

$$V_{j,t}^{e,\sigma}(s_t, \varepsilon_{j,t}) = \max_{d_{j,t}} \left\{ U_{j,t}^e(d_{j,t}, \varepsilon_{j,t}) + \mathbb{I}(d_{j,t} = \text{Enter}) \cdot \beta \mathbb{E} V_{j,t+1}^\sigma(x_{j,t+1}, s_{t+1}, \varepsilon_{j,t+1}) \right\}. \quad (8)$$

The expectations in Equations (7) and (8) are taken over other creators' actions in current period, values of the next period individual- and contest-state variables, as well as the next period private shocks.

<sup>8</sup>That is, potential entrant  $j$ 's decision depends only on the current contest-level state  $s_t$  and her current private utility shock  $\varepsilon_{j,t}$ ; incumbent  $i$ 's behavior depends only on the current contest-level state  $s_t$ , her current own state  $x_{i,t}$  and her current private utility shock  $\epsilon_{i,t}$ .

### 4.3. Equilibrium Concept

We solve this finite-horizon dynamic discrete game with private information for a Markov Perfect Equilibrium (MPE) in type-symmetric pure strategies. For the proposed structural model, a strategy  $\sigma^* = \{\rho^*, \lambda^*\}$  represents a MPE if, at any  $t$ , given everyone else is playing  $\sigma^*$ , an incumbent's and a potential entrant's best response are  $\rho^*$  and  $\lambda^*$ , respectively.

Following Milgrom and Weber (1985), we represent a MPE in probability space. We denote the *conditional choice probability (CCP)* corresponding to the MPE strategy  $\sigma^*$  as  $\mathbf{P}^*$ , in which the  $t^{\text{th}}$  element,  $P_t^*$ , characterizes creator strategies in the  $t^{\text{th}}$  period. Equilibrium probabilities are a fixed point  $\mathbf{P}^* = \mathbf{\Gamma}(\mathbf{P}^*)$ , where the function  $\mathbf{\Gamma}$  is the creators' *best response probability function*. Assuming both the incumbents' private shock ( $\epsilon_{i,t}$ ) and potential entrants' private shock ( $\epsilon_{j,t}$ ) follow the Type I extreme value distribution (Rust 1987), we then get that in equilibrium incumbent  $i$  follows

$$P_{i,t}^*(a_{i,t}|x_{i,t}, s_t) = \Gamma_{i,t}(a_{i,t}|x_{i,t}, s_t; \mathbf{P}_{-i}^*) = \frac{\exp(v_{i,t}^{\mathbf{P}^*}(x_{i,t}, s_t, a_{i,t}))}{\sum_{a'_{i,t}} \exp(v_{i,t}^{\mathbf{P}^*}(x_{i,t}, s_t, a'_{i,t}))}, \quad (9)$$

and potential entrant  $j$  follows

$$P_{j,t}^*(d_{j,t}|s_t) = \Gamma_{j,t}^e(d_{j,t}|s_t; \mathbf{P}_{-j}^*) = \frac{\exp(v_{j,t}^{e, \mathbf{P}^*}(s_t, d_{j,t}))}{\sum_{d'_{j,t}} \exp(v_{j,t}^{e, \mathbf{P}^*}(s_t, d'_{j,t}))}, \quad (10)$$

where  $v_{i,t}^{\mathbf{P}^*}$  and  $v_{j,t}^{e, \mathbf{P}^*}$  are incumbent and potential entrants' *choice specific value functions*:  $v_{i,t}^{\mathbf{P}^*}(x_{i,t}, s_t, a_{i,t}) = -c(a_{i,t}) + \beta \mathbb{E}[V_{i,t+1}^{\sigma^*}(x_{i,t+1}, s_{t+1}, \epsilon_{i,t+1})|a_{i,t}]$ ;  $v_{j,t}^{e, \mathbf{P}^*}(s_t, d_{j,t}) = -c^e(d_{j,t}) + \mathbb{I}(d_{j,t} = \text{Enter}) \cdot \beta \mathbb{E}[V_{j,t+1}^{\sigma^*}(x_{j,t+1}, s_{t+1}, \epsilon_{j,t+1})|d_{j,t}]$ .<sup>9</sup>

### 4.4. Winning Probability

The last element of our structural model to be explained is how creator  $i$ 's winning probability ( $Pr(i \text{ wins}|x_{i,T+1}, s_{T+1})$ ) in period  $T+1$  is calculated. Generally speaking, the creator who submits the highest quality design wins. So far, the quality of a design has been measured by its rating, which is observed in our data. One potential problem with this measure is that the observed ratings are truncated at 5-star – no matter how high a design's "true" quality is, the seeker can at most rate it as 5-star. If we neglect this problem, our model would fail to capture some important aspects of creator behavior observed in the data. For example, the model would predict that a creator whose current best rating is 5-star would have no incentive to make any additional submissions, as doing so only incurs cost but provides no benefit. However, in the data, we observe many cases where creators make additional submissions after receiving a 5-star rating; moreover, among creators who have received a 5-star rating, those who remain active in revising their design(s) or submitting new design(s) have a higher probability of winning the contest than those who become inactive.

<sup>9</sup> The derivation details are provided in Appendix B.

To resolve this problem and recover the underlying “true” quality (which we denote as  $x_{i,t}^{true}$ ) of a creator’s best design, we redefine the individual-level state as  $\tilde{x}_{i,t} = (x_{i,t}, \boldsymbol{\varpi}_{i,t})$ , with  $x_{i,t}$  still denoting the best rating that creator  $i$  has received, and an additional variable  $\boldsymbol{\varpi}_{i,t}$  recording the actions creator  $i$  has taken up to period  $t$  after receiving her first 5-star rating, if she has ever received any 5-star rating.  $\boldsymbol{\varpi}_{i,t}$  is a vector of three elements, respectively representing the number of times that the “redesign”, “revise” or “do-both” action has been taken by creator  $i$  since she received her first 5-star rating (excluding the action that results in the first 5-star rating). There are then two scenarios:

1. When  $x_{i,t} \in \{\text{NA}, 1, 2, 3, 4\}$ , i.e., creator  $i$ ’s current best score has not reached 5-star:  
 $\tilde{x}_{i,t} = (x_{i,t}, \boldsymbol{\varpi}_{i,t})$ , where  $\boldsymbol{\varpi}_{i,t} = (0, 0, 0)$ ;
2. When  $x_{i,\tau+1} = \dots = x_{i,t} = 5$ ,  $x_{i,\tau} < 5$ , i.e., creator  $i$  receives the first 5-star in period  $\tau + 1$ :  
 $\tilde{x}_{i,t} = (5, \boldsymbol{\varpi}_{i,t})$ , where  $\boldsymbol{\varpi}_{i,t} = (\boldsymbol{\varpi}_{i,t,\text{redesign}}, \boldsymbol{\varpi}_{i,t,\text{revise}}, \boldsymbol{\varpi}_{i,t,\text{do-both}})$ , with  $\boldsymbol{\varpi}_{i,t,a} = \sum_{l=\tau+1}^t \mathbb{I}(a_{i,l} = a)$ .

The contest-level state is then redefined as  $\tilde{s}_t$ , which summarizes the number of creators whose individual-level state is  $\tilde{x}_{i,t}$  for all possible values of  $\tilde{x}_{i,t}$ .  $\tilde{x}_{i,t}$  transitions in the following way:  $x_{i,t}$  still transitions in the same way as discussed in the previous subsection;  $\boldsymbol{\varpi}_{i,t}$  transitions deterministically as a function of the action creator  $i$  takes in period  $t$ :

$$\boldsymbol{\varpi}_{i,t+1,a} = \boldsymbol{\varpi}_{i,t,a} + \mathbb{I}(a_{i,t} = a), \text{ where } a \in \{\text{redesign}, \text{revise}, \text{do-both}\}. \quad (11)$$

With this newly-constructed state variable  $\tilde{x}_{i,t}$ , we can recover the true quality of a creator’s best design ( $x_{i,t}^{true}$ ). We assume that the first time creator  $i$  receives a 5-star rating, the true quality of that 5-star design is  $5 + \max(0, \xi_0)$ , where  $\max(0, \xi_0)$  is the part of quality that gets truncated by the integer rating of 5. After receiving a 5-star rating, every time a creator takes a non-null action  $a_{i,t} \in \{\text{redesign}, \text{revise}, \text{do-both}\}$ , there will be a corresponding continuous stochastic improvement ( $\max(0, \xi_{a_{i,t}})$ ). Hence, the mapping between a creator’s state variable ( $\tilde{x}_{i,t} = (x_{i,t}, \boldsymbol{\varpi}_{i,t})$ ) and the quality of her best design ( $x_{i,t}^{true}$ ) is as follows:

$$x_{i,t}^{true} = \begin{cases} x_{i,t} & \text{if } x_{i,t} < 5 \\ 5 + \max(0, \xi_0) + \sum_{a \in \{\text{redesign}, \text{revise}, \text{do-both}\}} \sum_{n=1}^{\boldsymbol{\varpi}_{i,t,a}} \max(0, \xi_a^{(n)}) & \text{if } x_{i,t} = 5, \end{cases} \quad (12)$$

where  $\max(0, \xi_a^{(n)})$  is the  $n^{\text{th}}$  realization of the stochastic improvement  $\xi_a$ .

We can now compute the probability of winning using the “true” quality in the terminal period  $T + 1$ . Let  $\bar{x}_{T+1}^{true} = \max_{i \in N_{T+1}} (x_{i,T+1}^{true})$  denote the maximum quality achieved by all creators in the game in period  $T + 1$ , then creator  $i$ ’s winning probability can be expressed as:

$$Pr(i \text{ wins} | \{x_{k,T+1}^{true}\}_{\forall k \in N_{T+1}}) = \mathbb{I}(x_{i,T+1}^{true} = \bar{x}_{T+1}^{true}) \frac{1}{\sum_{k \in N_{T+1}} \mathbb{I}(x_{k,T+1}^{true} = \bar{x}_{T+1}^{true})}. \quad (13)$$

Notice that Equation (13) implies two cases. (i) When  $\bar{x}_{T+1}^{true} > 5$ ,  $\bar{x}_{T+1}^{true}$  is a continuous variable. In this case, a tie is impossible – the designer of the highest-quality design wins the contest; (ii) when

$\bar{x}_{T+1}^{true} \leq 5$ , ties are possible. In case of ties, each creator whose best rating is  $\bar{x}_{T+1}^{true}$  wins the contest with an equal chance.<sup>10</sup> In the data, the majority of the contests fall into the case (i).

## 5. Estimation and Results

### 5.1. Estimation Strategy

Our approach to estimate the structural model proceeds in two steps. First, we estimate the parameters governing the state transition process, including the quality distribution of new entrants' submissions  $p^e(x_{j,t+1})$ , incumbents' quality transition probabilities  $p(x_{i,t+1}|x_{i,t}, a_{i,t}), \forall a_{i,t} \in \mathcal{A}$ , and the distributions that characterize improvements beyond 5-star, i.e.,  $\xi_0, \xi_{redesign}, \xi_{revise}$ , and  $\xi_{do-both}$ . In the second step, we embed the estimated state transition probabilities into the dynamic discrete game of creators' entry and follow-up actions, and estimate the parameters in creators' utility function. These two sets of parameters can be estimated separately under the conditional independence assumption.<sup>11</sup> Accordingly, the set of parameters in the structural model ( $\theta$ ) can be classified into two subsets: (1) parameters in creators' utility function ( $\theta_1$ ), and (2) parameters that govern the state transition process ( $\theta_2$ ).<sup>12</sup>

**5.1.1. Estimating the State Transition Process** We first estimate  $p^e(x_{j,t+1})$  and  $p(x_{i,t+1}|x_{i,t}, a_{i,t})$  using the frequency estimator, and use  $\theta_{21}$  to denote the parameters governing these transition probabilities.

Estimating the parameters governing the unobserved quality transition beyond 5-star is more challenging. The quality improvement beyond 5-star cannot be directly observed in the data; instead, it can only be inferred from the probability of winning. As discussed in the Model section, we introduce a series of random variables,  $\xi = \{\xi_0, \xi_{redesign}, \xi_{revise}, \xi_{do-both}\}$ , such that  $5 + \max(\xi_0, 0)$  represents the "true" quality of a creator's first 5-star submission, and  $\max(\xi_a, 0)$  represents the

<sup>10</sup> Here we assume in cases of ties, each creator has an equal chance, rather than each submission has an equal chance, to win the contest. The reason is that, when a creator decides to revise her existing submission(s), she typically picks her best performing style to revise; when a creator decides to redesign, she typically creates a new design style and abandons old design styles if this new design style is rated higher. Therefore, in period  $T + 1$ , we effectively assume that each creator has one design, i.e., the best among her submissions, in the seeker's consideration set.

<sup>11</sup> Under the conditional independence assumption (i.e.,  $p(x_{i,t+1}|x_{i,t}, a_{i,t})$  and  $\xi$ s are independent of  $p_\epsilon(\epsilon_{i,t})$ , and  $p^e(x_{j,t+1})$  is independent of  $p_\epsilon(\epsilon_{i,t})$ ), the transition probability functions  $p(x_{i,t+1}|x_{i,t}, a_{i,t})$  and  $p^e(x_{j,t+1})$  can be estimated from the transition data using a standard maximum likelihood method without solving the model.

<sup>12</sup> In the main estimation, we set  $\beta = 0.9$ , the number of potential entrants  $|M_t| = 300, \forall t$  to get results in Table 6.  $\beta$ 's identification is known to be impractical (Rust 1987), so we do not intend to estimate the discount factor. For the number of potential entrants, it cannot be identified together with entry costs. Therefore, we fix it to be the number of total active creators on the platform at any given time – the average is around 300. Subsequently, we conduct sensitivity analysis on both  $\beta$  and  $|M_t|$ , and show that the nature of our estimation results do not change.

quality improvement beyond 5-star resulting from action  $a \in \{redesign, revise, do-both\}$  (Equation (12)). We further assume that  $\xi$  follows a normal distribution,  $\xi \sim N(\mu, \sigma^2)$ . The parameter vector  $\{\mu_0, \sigma_0^2, \mu_{redesign}, \sigma_{redesign}^2, \mu_{revise}, \sigma_{revise}^2, \mu_{do-both}, \sigma_{do-both}^2\} := \theta_{22}$  can then be estimated by maximizing the following likelihood:

$$L_2(\theta_{22}) = \prod_{q=1}^Q \prod_{i=1}^{\|N_{q,T+1}\|} \left\{ Pr(i \text{ wins} | \{\tilde{x}_{k,T+1}\}_{k \in N_{q,T+1}}; \theta_{22})^{\mathbb{I}(i \text{ wins})} [1 - Pr(i \text{ wins} | \{\tilde{x}_{k,T+1}\}_{k \in N_{q,T+1}}; \theta_{22})]^{1 - \mathbb{I}(i \text{ wins})} \right\}, \quad (14)$$

where  $Q$  represents the number of contests in our estimation sample, and  $\|N_{q,T+1}\|$  is the total number of creators that submitted designs to contest  $q$ .

### 5.1.2. Estimating the Costs and Other Parameters in Creators' Utility Function

Once we obtain the consistent estimate of  $\theta_2 = \{\theta_{21}, \theta_{22}\}$ , denoted as  $\hat{\theta}_2$ , we can plug these estimated state transition probabilities into the dynamic discrete game model and solve the game for a MPE using backward induction. We use the Maximum Simulated Likelihood (MSL) method to estimate the parameters in creators' utility function ( $\theta_1$ ). Given a vector of candidate parameter values  $\theta_1$ , we numerically solve the dynamic game for the equilibrium strategy profile  $\sigma^*(\theta_1, \hat{\theta}_2)$  using a nested fixed-point approach. The MSL estimate  $\hat{\theta}_1$  is the vector that maximizes the likelihood of observing the actual choices in the data.

Finding the MPE for a contest with a large number of players involves solving a nested backward induction in an extremely large state space, which grows exponentially with the number of periods. To simplify the estimation and ensure the computational tractability of the policy simulations we divide each 7-day contest into three periods – Period 1 (Days 1-2), Period 2 (Days 3-5), and Period 3 (Days 6-7).<sup>13</sup> In a so-defined three-period contest, the seeker can give feedback at the end of the first and the second periods (the feedback after the last period will not be able to affect creator behaviors). Dividing contests into three periods significantly reduces the computational burden, but still allows us to capture not only the full feedback and the no feedback scenarios, but also the early feedback and the late feedback scenarios. However, the dimensionality of the joint distribution of  $(x_{i,t}, s_t)$  is still extremely large in a three-period contest.<sup>14</sup> To further reduce the computational burden, we adopt Keane and Wolpin's simulation and interpolation method (Keane and Wolpin

<sup>13</sup> We also try two alternative ways of dividing a contest into three periods: (i) Period 1 = Days 1-3, Period 2 = Days 4-5, and Period 3 = Days 6-7; (ii) Period 1 = Days 1-2, Period 2 = Days 3-4, and Period 3 = Days 5-7. The observed conditional choice probabilities and choice-specific state transition probabilities exhibit little sensitivity to how we define the periods. Details are available upon request.

<sup>14</sup> For example, the number of possible combinations of  $s_{t=3}$  for a three-period contest with 30 participants is  $1.15 \times 10^{11}$ , the different combinations of dividing 30 creators into 15 types  $\binom{30+15-1}{15-1}!$ .

1994, see Appendix D for details). The implied optimal choice probabilities from the nested fixed-point algorithm are  $Pr(d_{jt}|s_t; \theta_1, \hat{\theta}_2)$  ( $d_{jt} \in \mathcal{D}$ ) for new entrant  $j$  and  $Pr(a_{it}|x_{it}, s_t; \theta_1, \hat{\theta}_2)$  ( $a_{it} \in \mathcal{A}$ ) for incumbent creator  $i$ . The likelihood of observing  $m_{qt}$  out of  $\|M_{qt}\|$  potential entrants joining the contest  $q$  in period  $t$  is:<sup>15</sup>

$$f^e(m_{qt}|s_{qt}; \theta_1, \hat{\theta}_2) = \binom{\|M_{qt}\|}{m_{qt}} Pr(d_{qt} = enter|s_{qt}; \theta_1, \hat{\theta}_2)^{m_{qt}} (1 - Pr(d_{qt} = enter|s_{qt}; \theta_1, \hat{\theta}_2))^{\|M_{qt}\| - m_{qt}}, \quad (15)$$

and the likelihood of an incumbent  $i$  choosing follow-up action  $a_{qit} \in \mathcal{A}$  in period  $t$  of contest  $q$  is:

$$f(a_{qit}|x_{qit}, s_{qt}; \theta_1, \hat{\theta}_2) = \prod_{a \in \mathcal{A}} Pr(a_{qit} = a|x_{qit}, s_{qt}; \theta_1, \hat{\theta}_2)^{I(a_{qit}=a)}. \quad (16)$$

The joint likelihood to be maximized is then:

$$L_1(\theta_1, \hat{\theta}_2) = \prod_{q=1}^Q \prod_{t=1}^T \left[ f^e(m_{qt}|s_{qt}; \theta_1, \hat{\theta}_2) \cdot \prod_{i=1}^{\|N_{qt}\|} f(a_{qit}|x_{qit}, s_{qt}; \theta_1, \hat{\theta}_2) \right], \quad (17)$$

where  $\|N_{qt}\|$  is the number of incumbents in contest  $q$  up to period  $t$ . The maximum simulated likelihood estimator of  $\theta_1$  is

$$\hat{\theta}_1 = \arg \max_{\theta_1} \log L_1(\theta_1, \hat{\theta}_2). \quad (18)$$

In terms of the sources of identification, the state transition probabilities for ratings below 5-star and the distribution of new entrants' ratings ( $\theta_{21}$ ) are identified directly from the frequencies observed in the data.  $\theta_{22}$  is identified from the variation in the winning probability resulting from different combinations of follow-up actions taken by creators after receiving their first 5-star rating. These state transition processes, together with the observed entry and follow-up action choices in the panel data of creator activities, constitute the inputs for identifying the utility parameters. For any given contest-state, the observed fractions of incumbents taking different follow-up actions and the observed entry numbers will pin down the MSL estimators for all parameters in creators' utility functions ( $\theta_1$ ). For example, given the state transition probabilities, a large revision cost will reduce the predicted number of incumbents taking the revision action. If in the data, a small fraction of incumbents choose the revision action, then the estimated cost of the revision will be relatively large.

<sup>15</sup> We cannot observe each potential new entrant  $j$ 's decision; rather, we can only observe the number of total joining creators.

## 5.2. Results

**5.2.1. Estimates of State Transition Probabilities.** As discussed in Section 5.1, we use the frequency estimator to estimate the quality distribution of new entrants’ first submission(s) and the action-specific state transition probabilities among the observed states, that is, along the 5-star rating scale. The estimation results for  $p^e(x_{j,t+1})$  and  $p(x_{i,t+1}|x_{i,t}, a_{i,t})$  are presented in detail in Appendix D to save space in the body. Here, we highlight a few interesting findings: *do-both* is the most effective in improving individuals’ highest design rating, leading to on average larger improvements than *redesign* and *revision* do; however, the variation of those improvements is also high. For creators with low ratings (below or equal to 3-Star), *redesign* results in on average bigger, but more variable improvements than *revision*. For creators who have already received relatively high ratings (4-Star and 5-Star), *revision* leads to on average slightly larger improvements than *redesign* (Table 4).

**Table 4 Mean and Standard Deviation of Observed Rating Improvements, i.e., NA, 1- to 5-Star**

	Redesign		Revision		Do-both	
	Mean	S.t.d	Mean	S.t.d	Mean	S.t.d
Low Base Rating (NA, 1- to 3-Star)	2.16	3.28	1.70	2.62	3.08	4.39
High Base Rating (4- and 5-Star)	0.13	2.00	0.14	1.58	0.22	2.72

The estimation results for  $\theta_{22}$ , the set of parameters governing the distribution of the unobserved quality improvements beyond 5-star, are displayed in Table 5. Notice that we fix  $\xi_0$  to follow the standard normal distribution to achieve identification for other elements of  $\theta_{22}$ .<sup>16</sup> The estimation results suggest that after creators receive at least one 5-star rating, exploratory actions (*redesign*) are less likely to bring positive quality improvements, but the variance of these quality improvements is larger; by contrast, exploitative actions (*revise*) are more likely to bring positive quality improvements, but the variance in these quality improvements is smaller. Combining both exploratory and exploitative actions, *do-both* leads to a medium-level chance of positive quality improvements, but the variance of these quality improvements is very large.

**5.2.2. Estimation Results for Parameters in Creators’ Utility Function** Table 6 reports the estimation results for the parameters in creator’s utility function. The estimates are generally consistent with our intuition and statistically significant. Among incumbents’ follow-up actions, the estimated cost associated with the *redesign* action is higher than the estimated cost associated with the *revise* action. Also, the cost of the *do-both* action is higher than the cost of

<sup>16</sup> We are not able to simultaneously identify the distributions for  $\xi_0, \xi_{redesign}, \xi_{revise}, \xi_{do-both}$ , since shifting all  $\xi_a$  distributions horizontally or making them flatter/thinner simultaneously will not affect how we rationalize the winning realizations observed in the data.

**Table 5** Estimates for  $\theta_{22}$ , Parameters Governing Quality Improvements beyond 5-Star

Mean of $\xi_a$	Estimate	S.t.d of $\xi_a$	Estimate
$\mu(\xi_0)$	Fixed to 0	$\sigma(\xi_0)$	Fixed to 1
$\mu(\xi_{redesign})$	-0.052 (0.008)	$\sigma(\xi_{redesign})$	1.351 (0.003)
$\mu(\xi_{revise})$	0.230 (0.006)	$\sigma(\xi_{revise})$	0.517 (0.008)
$\mu(\xi_{do-both})$	0.022 (0.005)	$\sigma(\xi_{do-both})$	1.426 (0.010)

*Note:* The numbers in parenthesis are standard errors. Recall in Equation (12),  $max(\xi_a, 0)$  represents the quality improvement beyond 5-star resulting from action  $a$ .

*redesign* only or *revise* only, but lower than the sum of doing both separately.<sup>17</sup> An interesting aspect of these estimation results is that they are inferred based on observed actions. Therefore, one may wonder if more direct measures would also lead to qualitatively similar findings. Of course, we cannot directly observe effort, but one proxy we can observe is the amount of time that elapses between two consecutive submissions. We find that the time for a *redesign* submission (11.35 hours) is longer than the time for a *revision* submission (9.22 hours), supporting our inferred estimation result that *redesigns* are more costly than *revisions*.

In terms of the estimates for the period-specific entry costs, we have two observations. First, the estimated entry costs are always higher than costs associated with incumbents' follow-up actions. This is expected, because in addition to the effort required to submit a design, the entrant also spends effort discovering the contest, understanding the problem specification, etc. Second, the entry cost is decreasing over time. One plausible explanation is that, by default, contests that are closing soon rank higher on the site's contest list. A higher position makes it easier for creators to discover the contest, which reduces one component of the entry cost discussed above.

The marginal utility of money ( $\alpha$ ), or the number of utils a creator receives from getting an additional dollar of award, is estimated to be 0.034. This implies that a \$200 award corresponds to around 6.8 utils, which is about 2.4 times the cost of *redesign*, 3.1 times the cost of *revision*, and 2.0 times the cost of *doing-both*. Comparing with the utility from rewards, the estimated costs might appear relatively high at first glance, but this is consistent with the rather small number of entries (the average number of participants is 29 for \$200 contests) and infrequent follow-up submissions (one creator only submits on average 0.957 revision submissions and 0.594 redesign submissions in a contest). We can also convert the costs measured in utils to the corresponding dollar amounts. For example, the monetary cost of submitting one or more new designs is  $2.790/0.034 = \$83$ , the cost of making revision(s) is \$65; the cost of doing both is \$98; and the period 1-3 entry costs are \$147, \$128, and \$107 respectively. The estimated structural model also predicts that creators who have received high ratings (4-star and 5-star) are almost twice as likely to make follow-up

<sup>17</sup> Notice that these cost numbers are not "per-submission costs"; rather, the costs should be interpreted as the cost of employing only the exploration strategy, only the exploitation strategy, and both strategies in one period. Each strategy can possibly involve multiple submissions.



submissions, compared to those whose best rating is 3-star or lower (38% versus 21%). If creators indeed take follow-up actions, those who have received low ratings (3-star or lower) are about fifty percent more likely to choose *redesign*, compared to those whose best rating is 4- or 5-star (30% versus 20%). These results indicate that high performers prefer the exploitative strategy to the exploratory strategy; low performers tend to make fewer follow-up submissions, and are more likely to choose the exploratory strategy when doing so.

**Table 6** Estimates for  $\theta_1$ , Parameters in Creators' Utility Function

Parameter	Description	Estimate	Standard Error
$\alpha$	Marginal Utility of Money	0.034	(0.002)
$c(a_{i,t} = \textit{redesign})$	Cost of Redesign	2.790	(0.010)
$c(a_{i,t} = \textit{revise})$	Cost of Revision	2.205	(0.021)
$c(a_{i,t} = \textit{do-both})$	Cost of Doing Both	3.319	(0.031)
$c_1^e(d_{i,t} = \textit{enter})$	Period 1 Entry Cost	4.958	(0.028)
$c_2^e(d_{i,t} = \textit{enter})$	Period 2 Entry Cost	4.326	(0.043)
$c_3^e(d_{i,t} = \textit{enter})$	Period 3 Entry Cost	3.599	(0.014)

*Note:* The numbers in parenthesis are standard errors, calculated with bootstrapping.

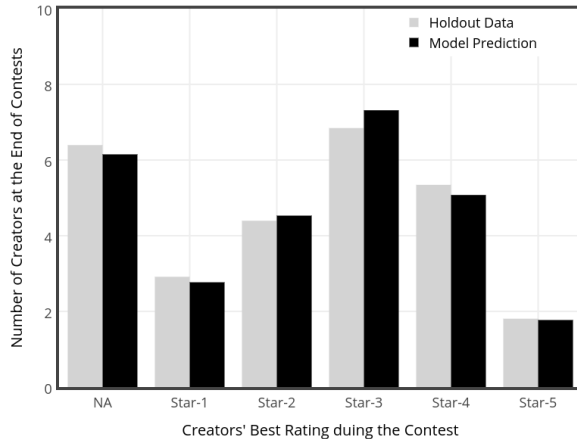
### 5.3. Model Performance

To test the performance of our structural model in predicting contest outcomes, we consider a hold-out sample of 418 contests and 11,578 contest-creator combinations. The contests in the hold-out sample are similar to those in the sample used for the estimation in terms of award level, and are all buyer-assured, single-winner, public and 7 days long. Additionally, the seeker in all the 418 contests provided feedback throughout the contest horizon.

We simulate the competition dynamics for the 418 contests using the estimated structural model. For each contest, we draw 50,000 sample paths, and calculate the average numbers of creators whose highest rating at the end of the contest is NA, 1-star, ..., and 5-star, respectively, across the 50,000 simulations, as the “simulated” numbers of NA, 1-star, ..., and 5-star creators for that contest. In Figure 2, black columns represent the average simulated number of creators whose best rating is NA, 1-star, ..., or 5-star across the 418 contests. The grey columns represent the actual average numbers of creators whose best rating is NA, 1-star, ..., and 5-star across the 418 contests observed in the data. Figure 2 shows that the actual numbers observed in the data closely match the simulated numbers. Therefore, we conclude that our model performs well in predicting contest outcomes in the holdout sample.

### 5.4. Robustness Checks

We designed our main model to be parsimonious but effective. To ensure that our empirical results are robust to our modeling choices and assumptions, we conduct a series of robustness checks.



**Figure 2 Model Performance**

**5.4.1. Heterogeneity** In the main model, we assume creators are ex-ante homogeneous; that is, the heterogeneity among creators is captured by their realized ratings; before the ratings of their submissions are disclosed, they are homogeneous. We believe that this is a reasonable assumption in our setting for two reasons. First, the logo design contest is likely to be a so-called “ideation project” (Terwiesch and Xu 2008), where the impact of participants’ endowed expertise is attenuated by the fact that the notion of quality is highly subjective – it is based on seekers’ private tastes rather than objective quality measures, as quoted from a popular platform: “These ratings are subjective. A star rating doesn’t reflect your design skill – it indicates the personal preferences of the contest holder.” Hence, a creator who performs well in other design contests, or has participated in a large number of contests does not necessarily have an advantage in a new contest she chooses to participate in. Second, even though the creator population on the platform is highly diverse in their experience level and background, those who frequently participate in contests and thus contribute more to the model estimation are relatively homogeneous.

To ensure that our results are not sensitive to the homogeneity assumption, we perform the following robustness check, which accounts for the possibility of creator ability/experience heterogeneity. We use the *Reputation Score* (on the scale of 0 – 100) the platform assigns to creators to measure their ability/experience level. This *Reputation Score* is computed by the platform and is displayed on every creator’s profile page, and it summarizes the ratings of the creator’s past submissions, her level of participation, her history on the platform, and her community behavior (e.g., frequencies of visiting the site, reporting problems, and participating in the forums). If a creator’s *Reputation Score* is above or equal to 70, she is classified into the high type (H); otherwise, she is classified into the low type (L).<sup>18</sup> Based on this classification, we (1) conduct regression

<sup>18</sup> We use 70 as the cutoff score, because 70 is the starting-point score assigned to a new creator who just joined the platform, from where the system adjusts the creator’s score upwards or downwards according to her performance and activity level.

analyses to test whether there is strong evidence for heterogeneous behavior between the high- and low-types of creators; and (2) re-estimate our structural model using stratified samples and show the robustness of the estimation results.

**Regression Analysis** We first regress the percentage of the high-type creators among all creators joining contests on day  $t$  ( $\% \Delta H_{qt} = \frac{\Delta(\text{No. High Type Creators})_{qt+1}}{\Delta(\text{No. Creators})_{qt+1}}$ )<sup>19</sup> on  $\mathbf{W}_{qt}$  and  $\mathbf{Y}_{qt}$ , to test whether the high-type creators are more/less likely to join contests that are more/less competitive. The estimation results of this regression suggest that neither the number of existing submissions with high ratings, nor the number of existing submissions with low ratings has a significant effect on  $\% \Delta H_{qt}$ , and that the model is not significant with  $F\text{-Stats} = 1.244$  ( $df = 7; n = 4793$ ).<sup>20</sup> In other words, there are not disproportionately high- or low-type creators joining a contest when more high or low ratings are disclosed. Therefore, there is little evidence for any heterogeneous entry behavior between the high-type and the low-type creators.

Next, we test whether the high-type and low-type creators differ in their decisions on follow-up actions by estimating a variation of Equation (2), in which two additional independent variables – the focal creator’s type dummy ( $I(H)_i$ )<sup>21</sup> and the percentage of high-type creators among all existing creators in contest  $q$  on day  $t - 1$  ( $H_{qt-1} = \frac{(\text{No. High Type Creators})_{qt+1}}{(\text{No. Creators})_{qt+1}}$ ). The results of this multinomial regression suggest that  $I(H)_i$  is not significantly correlated with the probabilities of *revision* and *do-both*, and is only marginally significantly correlated with *redesign*;  $\% H_{qt-1}$  is not significantly correlated with any of the follow-up actions. This indicates that neither the focal incumbent creator  $i$ ’s type nor the percentage of high-type creators in the contest significantly affects creator  $i$ ’s choice of follow-up actions, after controlling for the individual-level and contest-level state variables. This finding, along with the results for the previous regression, supports our argument that creators’ participation behavior is not significantly affected by either their own ability/experience, or that of their rivals. The complete regression results can be found in Appendix E.

**Stratified Analysis** To further demonstrate that our estimation results are robust to the inclusion/exclusion of creator ex-ante heterogeneity in their ability/experience, and that not incorporating creator ex-ante heterogeneity does not affect the nature of our estimation results, we re-estimate the main model using stratified sub-samples. Operationally, we stratify contests in the complete sample into *High-Type Concentrated Contests*, *Low-Type Concentrated Contests*, and *Balanced Contests*, based on the percentage of high-type creators in the contest,<sup>22</sup> and we estimate our

<sup>19</sup> We added 1 to both the numerator and the denominator to avoid un-defined numbers.

<sup>20</sup> Among all independent variables, the only significant one is  $(\text{No. Submissions})_{qt-1}$ : it is only marginally significant ( $p\text{-value} = 0.028$ ) with a small magnitude (0.001).

<sup>21</sup>  $I(H)_i = 1$ , if  $\text{ReputationScore}_i \geq 70$ ; otherwise  $I(H)_i = 0$ .

<sup>22</sup> *High-Type Concentrated Contests*: contests with a high proportion ( $\geq$  upper quartile of the proportion observed in all contests in the data) of high ability creators; *Low-Type Concentrated Contests*: contests with a low proportion ( $\leq$

structural model using *High-Type Concentrated* and *Low-Type Concentrated Contests* separately, and compare the estimation results based on the two sub-samples with the full-sample results.

**Table 7 Results for Stratifies Analysis and the Model with Non-Monetary Incentives**

Parameter	Main Model		High-Type Concentrated		Low-Type Concentrated		With Non-Monetary Incentive	
$\alpha$	0.034	(0.002)	0.034	(0.001)	0.033	(0.001)	0.028	(0.001)
$c(\text{redesign})$	2.790	(0.010)	2.790	(0.044)	2.790	(0.011)	2.784	(0.008)
$c(\text{revise})$	2.205	(0.021)	2.157	(0.027)	2.254	(0.035)	2.204	(0.006)
$c(\text{do-both})$	3.319	(0.031)	3.313	(0.023)	3.324	(0.023)	3.293	(0.010)
$c_1^e(\text{enter})$	4.958	(0.028)	4.919	(0.030)	5.016	(0.053)	5.032	(0.097)
$c_2^e(\text{enter})$	4.326	(0.043)	4.388	(0.016)	4.411	(0.014)	4.466	(0.025)
$c_3^e(\text{enter})$	3.599	(0.014)	3.814	(0.032)	3.490	(0.020)	3.814	(0.020)
$R_{nm}$	–		–		–		0.233	(0.007)

*Note:* The numbers in parenthesis are standard errors.

As we can see in Table 7, the *High-Type Concentrated Contests* and *Low-Type Concentrated* sub-samples yield similar structural estimates, and both sets of estimates are very similar to our main estimation results. We conclude that creator ability/experience heterogeneity does not meaningfully affect our estimation results.

**5.4.2. Including Non-Monetary Incentives** Another assumption we make in our main model is that the positive utility a creator receives in the terminal period only comes from the financial reward given by the seeker (see Equation 3). One might argue that creators can get non-monetary rewards (e.g., learning by participating, pure joy of designing, and building up design portfolios) from participating in these design contests as well. To see whether our main estimation results are robust to the inclusion of non-monetary incentives, we revise the utility function for the rewarding period ( $T + 1$ ) as follows.

$$U_{i,T+1}(x_{i,T+1}, s_{T+1}) = \alpha R \cdot Pr(i \text{ wins} | x_{i,T+1}, s_{T+1}) + R_{nm}, \quad (19)$$

where  $R_{nm}$  is the additional non-monetary reward creator  $i$  receives in the terminal period. Note that the financial reward  $R$  is only received when a creator wins the contest, while any creator receives the non-monetary reward  $R_{nm}$  as long as she participates. The estimation results of the revised model are reported in the last column of Table 7. As we can see, the inclusion of the non-monetary incentives has little effect on the estimates of the utility parameters.

lower quartile of the proportion observed in all contests in the data) of high ability creators; and *Balanced Contests*: the remaining contests.

**5.4.3. Other Robustness Checks** We further test the sensitivity of the estimation results with respect to the discount factor  $\beta$ , the number of potential entrants  $\|M_t\|$ , and the SIFT cutoff, and summarize the results in Table 8. Not surprisingly, most of the cost estimates increase with  $\beta$ , because a higher  $\beta$  increases the expected utility from future periods, and hence the model needs larger cost estimates to rationalize the observed patterns of *entry*, *redesign*, *revision*, and *do-both*. Additionally, entry costs for all periods increase with the assumed number of potential entrants, as models that assume more potential entrants need higher entry costs to rationalize the number of entrants observed in the data. Lastly, the cost of *revision* increases and the cost of *redesign* decreases as we tune up the SIFT cutoff; since higher SIFT cutoff categorizes more actions into *redesign*, the model needs a lower *redesign* cost estimate and a larger *revision* cost estimate to rationalize the observed patterns in creators’ follow-up actions. Overall, the qualitative nature of the results are the same under different assumptions for the discount factor, the number of potential entrants, and the SIFT cutoff.

**Table 8 Sensitivity Analysis with respect to  $\beta$ ,  $\|M_t\|$ , and SIFT cutoff**

	<b>Main</b>	<b><math>\beta=0.95</math></b>	<b><math>\beta=0.85</math></b>	<b><math>\ M_t\ =350</math></b>	<b><math>\ M_t\ =250</math></b>	<b>SIFT 0.35</b>	<b>SIFT 0.45</b>
$\ M_t\ $	300	300	300	350	250	300	300
$\beta$	0.9	0.95	0.85	0.9	0.9	0.9	0.9
SIFT cutoff	0.4	0.4	0.4	0.4	0.4	0.35	0.45
$\alpha$	0.034 (0.002)	0.030 (0.001)	0.034 (0.002)	0.033 (0.003)	0.030 (0.001)	0.031 (0.002)	0.029 (0.001)
$c(\textit{redesign})$	2.790 (0.010)	2.792 (0.009)	2.790 (0.008)	2.788 (0.006)	2.789 (0.008)	2.997 (0.012)	2.590 (0.009)
$c(\textit{revise})$	2.205 (0.021)	2.205 (0.011)	2.204 (0.006)	2.205 (0.011)	2.212 (0.015)	2.047 (0.031)	2.217 (0.052)
$c(\textit{do-both})$	3.319 (0.031)	3.298 (0.025)	3.296 (0.013)	3.297 (0.015)	3.294 (0.027)	3.476 (0.183)	3.267 (0.059)
$c_1^e(\textit{enter})$	4.958 (0.028)	5.070 (0.023)	4.811 (0.018)	5.100 (0.029)	4.736 (0.017)	4.932 (0.015)	4.933 (0.022)
$c_2^e(\textit{enter})$	4.326 (0.043)	4.292 (0.023)	4.270 (0.025)	4.483 (0.010)	4.148 (0.019)	4.291 (0.025)	4.245 (0.026)
$c_3^e(\textit{enter})$	3.599 (0.014)	3.578 (0.027)	3.593 (0.019)	3.722 (0.013)	3.407 (0.020)	3.571 (0.022)	3.555 (0.033)

*Note:* The numbers in parenthesis are standard errors.

## 6. Counterfactual Simulations

Most crowdsourcing platforms suggest that seekers provide feedback throughout the contest horizon. One of these platforms even has the mantra that “give feedback early - give feedback often”. Following the platform’s suggestion, more than 70% of contest seekers provide feedback throughout the contest horizon in our data.

To assess this common practice and quantify the impact of feedback on the contest outcome, in this section, we use the estimation results of the structural model in Section 5.2 to conduct

policy simulations. Specifically, we experiment with four alternative feedback disclosure policies: (1) the seeker provides feedback throughout the contest (*full feedback*), (2) the seeker does not provide feedback at all (*no feedback*), (3) the seeker provides feedback only in the first half of the contest (*early feedback*), and (4) the seeker provides feedback only in the second half of the contest (*late feedback*). We compare the policy performance using the following three main metrics: (1) the quality of the best design, (2) the number of top performers (creators with at least one 5-star design), and (3) the total number of creators participating in the contest. To capture the possibility that the relative performance of the four alternative policies may vary across different award levels, we conduct the simulation at a series of award levels ranging from \$200 to \$400, with steps of \$50. The award levels of most contests on the platform fall in this range.

We now detail the four models, one for each of the policies (full feedback, no feedback, early feedback, and late feedback). The models differ in the feedback scheme the seeker commits to and follows. As a result, the creators' information set is different under different policies.<sup>23</sup> We assume that creators know which feedback policy is currently used and are forward-looking when making their decisions. Below is a summary of creators' information structure under the four feedback policies:

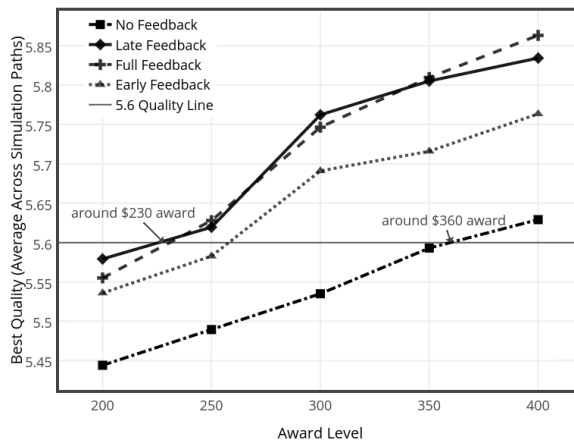
1. **Full feedback:** Existing creators and potential entrants observe the revealed ratings, and thus base the calculation of their chance of winning the contest on the revealed ratings.
2. **No feedback:** Existing creators and potential entrants only get to see prior actions taken by existing creators, but not the realized ratings. However, they can infer their winning chance from the observed actions, as the actions taken by creators will affect the distribution of the best quality of their submissions, which in turn affects creators' winning probabilities. Therefore, creators will base their decisions on the available action information, including their own actions and other contest participants' actions.
3. **Early feedback:** The competition structure under the early feedback policy is the same as that under the full feedback policy until the feedback is turned off in the second period. When making decisions in the third period, creators can only observe ratings disclosed at the end of the first period and participants' actions in the second period. Hence, they have to infer the quality of their own and opponents' submissions from the first-period realized ratings and the second-period actions.
4. **Late feedback:** Before the feedback disclosure is turned on, creators only observe their own and other contest participants' actions, but not the realized ratings; hence in this period of time, their decisions are based on their own and other contest participants' prior actions,

<sup>23</sup> Correspondingly, the way we define state variables and state transitions differs with different feedback policies.

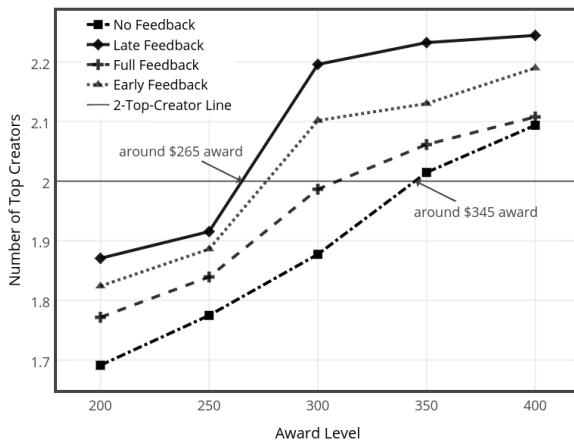
similar to the no feedback case. Once the seeker discloses the performance feedback at the end of the second period, the information structure of the competition onward becomes the same as that in the full feedback case.

In essence, the four feedback disclosure policies correspond to four types of dynamic games, each of which has its unique information structure. Therefore, our counterfactual experiments involve solving four different types of dynamic games. At each reward level, we use numerical methods to solve for the equilibrium for each of these games, because none of them has a closed-form solution. Given an award level and a feedback policy, we simulate 50,000 independent contest outcomes, and report the average value of the performance metrics across the 50,000 simulation paths.

(a) Maximum Quality



(b) The Number of Top Performers



(c) The Number of Participants

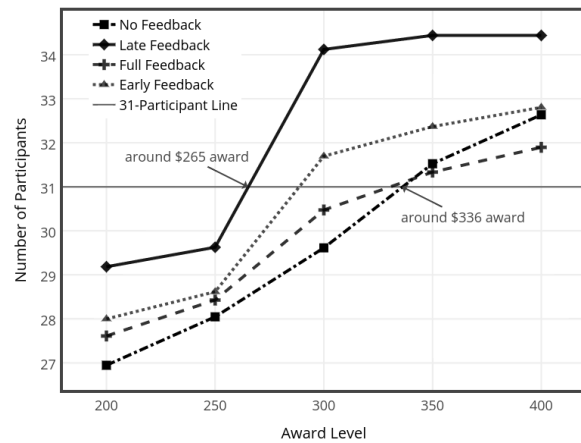


Figure 3 Contest Outcome Metrics under the Alternative Feedback Policies

### 6.1. Performance Metric I – The Maximum Quality

We first consider the best quality achieved by all contest participants. As discussed earlier, in contests with a single winner, the seeker is likely to care most about the quality of the best design (the extreme value). In Figure 3a, we can see that the no feedback policy is dominated by all the other three policies, confirming that giving feedback generally improves the contest outcome in terms of the maximum quality.

To further visualize the value of providing feedback, we can set a target best quality, and compare how much monetary incentive (award) the seeker should provide to achieve the pre-specified target. For example, if a seeker wants to achieve a maximum quality of 5.6 (the horizontal line in Figure 3a), he can either (1) set the award at approximately \$360 and provide no feedback, or (2) set the award at around \$230 and provide feedback throughout the contest, or only in the second half of the contest. In this example, giving feedback throughout or only in the second half of the contest can save the seeker \$130, which is roughly one third of the award that the seeker has to pay if he decides not to provide performance feedback at all.

However, should performance feedback be disclosed “as early as possible and as frequently as possible” as suggested by crowdsourcing platforms? Not necessarily. Although the full feedback policy performs best most of the time, our simulation results also show that the late feedback policy performs as well as the full feedback policy at nearly all award levels experimented with, and its performance even exceeds the full feedback policy when the award is around \$200 or \$300.

### 6.2. Performance Metric II – The Number of Top Performers

In many cases, innovation seekers would also like to have a large number of top creators, because more top performers can provide a richer set of high-quality submissions. Therefore, we consider the number of top creators, i.e., the number of creators who achieve high ratings (5-star), as the second performance metric. The performance of the four feedback policies in this metric is summarized in Figure 3b. It turns out that the late feedback policy performs best at all award levels experimented with. This may sound counter-intuitive at first, but when we look closer at the mechanism through which feedback affects creators’ activities, the result makes sense. Under the full feedback policy, feedback is provided from early on when there are only a handful of participants. Most incumbents (except the very few in the lead) are discouraged from taking follow-up actions, and potential entrants are also discouraged from entering the contest. In contrast, under the late feedback policy, the feedback is muted at the early stage of the contest – no creator is revealed to be in an advantageous position. More existing participants will remain active in making new submissions, and more potential entrants will be willing to join the contest. By the time when the performance feedback is disclosed, there will be more creators having good-quality submissions,



and they will continue making submissions in the last period. Hence, under the late feedback policy, there will be more top performers at the end of the contest.

As we did for the first performance metric, we can visualize the value of providing feedback on the second performance metric (the number of top performers). For example, if a seeker wants to obtain an average of two top creators, he has four options: (1) awarding around \$265 and providing late feedback, (2) awarding around \$273 and providing early feedback, (3) awarding around \$305 and providing full feedback, or (4) awarding around \$345 and providing no feedback. In this case, by giving feedback in the second half of the contest, the seeker saves almost a quarter of the award, compared to not providing feedback at all.

### **6.3. Performance Metric III – The Number of Participants**

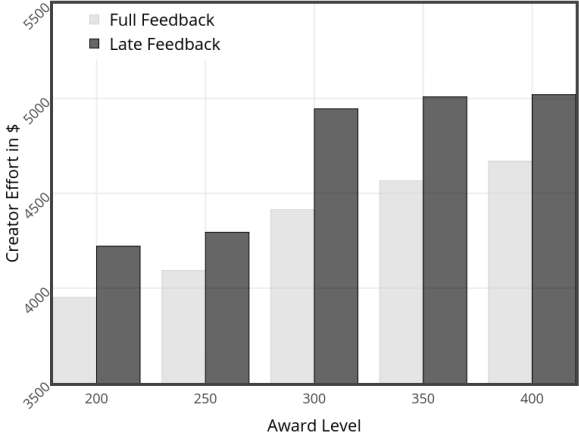
The results of our counterfactual simulations also suggest that the late feedback policy outperforms the other three policies in attracting creators to the contest at all award levels experimented with (Figure 3c). The reason for this is similar to the reason we provided in the previous subsection regarding the number of top performers – the disclosure of high ratings can discourage entry. However, one may ask, if the disclosure of the performance feedback always discourages the majority of the contest participants, why don't we always mute the performance feedback (choose the no feedback policy)? The issue with the no feedback policy is that, without the ability to distinguish between high-performers and low-performers, the expected probability for each person to win the contest decreases with the number of participants. Consequently, as the contest becomes more and more crowded, potential entrants are less willing to join the contest, and existing participants are less willing to take costly follow-up actions as well.

As the analysis for the previous two performance metrics, we can also quantify the value of releasing feedback with the objective of attracting more participants, e.g., see the horizontal line drawn halfway up the graph in Figure 3c.

### **6.4. Additional Findings**

In addition to our main results, our policy simulations also help us explore possible reasons why most platforms encourage seekers to give feedback throughout the contest period, which corresponds to the full feedback policy in our model. At the first sight, the platform suggestion contradicts with our finding that the late feedback policy performs well in all three matrices. Yet our metrics focused on the seeker's objectives, which may be somewhat different from the platform's objectives. Intuitively, the platform would like to run sustainably by wisely allocating creator efforts. Indeed, the following quotes admonish creators to conserve their effects: "if you find yourself frustrated by a project – withdraw and move on"; "focus your attention and efforts on the projects that you can get something out of". In other words, it is preferable to the platform if a good contest

performance can be achieved with creators spending less effort in total. By analyzing the policy simulation results, we can show that full feedback performs better in saving more creator efforts than the late feedback policy. Under the full feedback policy, the seeker strongly encourages top performers by disclosing their good performance, which, at the same time, saves lower-performers from wasting their efforts. In contrast, under the late feedback policy, the seeker remains silent about creator performance until the last stage, keeping incumbents generally active, and potential entrants more willing to join. Hence, total efforts from all participants are larger under the late feedback policy.



**Figure 4 The Sum of Creator Efforts in Each Contest (in \$)**

### 7. Discussion and Conclusion

Facilitated by technology, crowdsourcing contests are becoming an increasingly popular mechanism to source innovation from large-scale online markets. As the quality and quantity of innovations sourced through crowdsourcing contests highly depend on the design of such contests, one of the most crucial questions facing real-world innovation seekers is how to design effective contests to achieve better outcomes. In this paper, we empirically examine the important, albeit somewhat understudied element of crowdsourcing contest design: the role of performance feedback on the outcome of crowdsourcing contests. We develop a dynamic structural model to capture the economic processes that drive creators’ participation behavior, which highlights how existing contest participants and potential new entrants react to the disclosed performance feedback. The structural model explicitly considers potential entrants’ endogenous entry processes, and distinguishes between incumbent creators’ exploratory and exploitative follow-up actions. We recover the parameters in the structural model using a rich real-world data set on custom logo design contests collected from a major crowdsourcing platform. Our structural analysis yields insights into the mechanics

that drive creators' behavior under different feedback policies, as summarized on pages 3-4 of the Introduction.

Our policy simulations provide important insights into the role of feedback disclosure policy in crowdsourcing contests – they not only compare contest outcomes under alternative feedback disclosure policies, but also quantify the value of providing performance feedback on different performance metrics. In particular, we show that if all that the seeker cares about is the maximum quality achieved, both the full feedback policy and the late feedback policy perform quite well. If the seeker's objective is to maximize the number of high performers, or the total number of participants in the contest, the late feedback policy is the best option. Feedback helps guide creators' exploration and exploitation decisions, but can have a discouraging effect on entries and incumbents' follow-up actions. The late feedback policy attains the former benefit while mitigating the latter problem, by only giving feedback after many creators have had a chance to enter. If we further take into consideration the cost associated with monitoring submissions and providing performance feedback in real time, the late feedback policy becomes even more attractive. Given the above, our study may help seekers make better decisions about feedback policies in practice, by highlighting the merits of the late feedback policy.

As one of the first empirical studies of large-scale crowdsourcing contests, our paper makes a number of important contributions but also has limitations. First, in our current analysis, we assume the contests are independent of each other, and have not considered the creators' choices among concurrent on-going contests. A systematic analysis of what factors affect creators' choices of which contest to join could be a productive direction of future research. Second, we focus on the role of quantitative performance feedback (i.e., ratings), and have not considered qualitative feedback, given through private messages, for two reasons: (1) qualitative feedback in the form of private messages is inaccessible to us; and (2) based on an interview with a marketing manager of a major online crowdsourcing platform, qualitative feedback occurs much less frequently than quantitative performance feedback. However, in other settings where the qualitative feedback is more prevalent and available to researchers, the effects of qualitative feedback could be a fertile direction for future research.

Despite these limitations, our paper is the first to provide a comprehensive dynamic structural framework to analyze creators' behavior in crowdsourcing contests. With the use of the structural model, we are able to disentangle intertwining effects of feedback on the outcome of crowdsourcing contests, helping both practitioners and researchers obtain a more comprehensive understanding of this increasingly popular new approach for sourcing innovation. Our policy simulation results shed light on how to choose the optimal feedback disclosure policy, based on the objective the innovation seeker wants to achieve. In addition to advancing the managerial understanding of this important

issue, our work is also one of the first examples of using a structural modeling approach to study the design of crowdsourcing contests. Although the focus of our paper is the role of feedback, the structural framework we propose can be used to analyze other design issues in crowdsourcing contests. We hope that our work can pave the way for future research in this area.

## Acknowledgments

The authors gratefully acknowledge seminar participants at Georgia Institute of Technology and the University of Michigan for valuable feedback.

## References

- Aguirregabiria V, Mira P (2010) Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1):38–67.
- Aksin Z, Ata B, Emadi SM, Su CL (2013) Structural estimation of callers’ delay sensitivity in call centers. *Management Science* 59(12):2727–2746.
- Ales L, Cho SH, Körpeoğlu E (2016a) Innovation tournaments with multiple contributors. *Working Paper*.
- Ales L, Cho SH, Körpeoğlu E (2016b) Optimal award scheme in innovation tournaments. *Working Paper*.
- Allon G, Bassamboo A, Çil EB (2012) Large-scale service marketplaces: The role of the moderating firm. *Management Science* 58(10):1854–1872.
- Allon G, Bassamboo A, Çil EB (2014) A service marketplace with multiple classes and multiple skilled agents. *Working Paper*.
- Allon G, Federgruen A, Pierson M (2011) How much is a reduction of your customers’ wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management* 13(4):489–507.
- Aoyagi M (2010) Information feedback in a dynamic tournament. *Games and Economic Behavior* 70(2):242–260.
- Bassamboo A, Cui R, Moreno A (2015) The wisdom of crowds in operations: Forecasting using prediction markets. *Working Paper*.
- Bimpikis K, Ehsani S, Mostagir M (2014) Designing dynamic contests. *Working Paper*.
- Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science* 57(5):843–863.
- Boudreau KJ, Lakhani KR (2015) “open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy* 44(1):4–19.
- Boudreau KJ, Lakhani KR, Menietti M (2016) Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *The RAND Journal of Economics* 47(1):140–165.

- Bray RL, Mendelson H (2012) Information transmission and the bullwhip effect: An empirical investigation. *Management Science* 58(5):860–875.
- Bray RL, Mendelson H (2015) Production smoothing and the bullwhip effect. *Manufacturing & Service Operations Management* 17(2):208–220.
- Che YK, Gale I (2003) Optimal design of research contests. *The American Economic Review* 93(3):646–671.
- Dahan E, Mendelson H (2001) An extreme-value model of concept testing. *Management science* 47(1):102–116.
- Dechenaux E, Kovenock D, Sheremeta RM (2015) A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics* 18(4):609–669.
- Ederer F (2010) Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy* 19(3):733–769.
- Elmaghraby W (2007) Auctions within e-sourcing events. *Production and Operations Management* 16(4):409–422.
- Erat S, Krishnan V (2012) Managing delegated search over design spaces. *Management Science* 58(3):606–623.
- Fudenberg D, Gilbert R, Stiglitz J, Tirole J (1983) Preemption, leapfrogging and competition in patent races. *European Economic Review* 22(1):3–31.
- Fullerton RL, McAfee RP (1999) Auction entry into tournaments. *Journal of Political Economy* 107(3):573–605.
- Gershkov A, Perry M (2009) Tournaments with midterm reviews. *Games and Economic Behavior* 66(1):162–190.
- Girotra K, Terwiesch C, Ulrich KT (2010) Idea generation and the quality of the best idea. *Management Science* 56(4):591–605.
- Goltsman M, Mukherjee A (2011) Interim performance feedback in multistage tournaments: The optimality of partial disclosure. *Journal of Labor Economics* 29(2):229–265.
- Gross DP (2016a) Performance feedback in competitive product development. *Working Paper*.
- Gross DP (2016b) Creativity under fire: The effects of competition on creative production. *Working Paper*.
- Guajardo JA, Cohen MA, Netessine S (2015) Service competition and product quality in the US automobile industry. *Management Science* 62(7):1860–1877.
- Halac M, Kartik N, Liu Q (2016) Contests for experimentation. *Journal of Political Economy, Forthcoming*.
- Hu M, Li X, Shi M (2015) Product and pricing decisions in crowdfunding. *Marketing Science* 34(3):331–345.
- Hu M, Wang Z (2015) Secrecy versus transparency in sales of network goods. *Working Paper*.
- Hyndman K, Parmeter CF (2015) Efficiency or competition? A structural econometric analysis of canada’s aws auction and the set-aside provision. *Production and Operations Management* 24(5):821–839.

- Keane MP, Wolpin KI (1994) The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics* 648–672.
- Kim SW, Olivares M, Weintraub GY (2014) Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Science* 60(5):1180–1201.
- Kornish LJ, Ulrich KT (2011) Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Management Science* 57(1):107–128.
- Körpeoğlu E, Cho SH (2015) Incentives in contests with heterogeneous solvers. *Working Paper*.
- Li J, Granados N, Netessine S (2014) Are consumers strategic? Structural estimation from the air-travel industry. *Management Science* 60(9):2114–2137.
- Liu TX, Yang J, Adamic LA, Chen Y (2014) Crowdsourcing with all-pay auctions: A field experiment on taskcn. *Management Science* 60(8):2020–2037.
- Lowe DG (1999) Object recognition from local scale-invariant features. *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, 1150–1157 (Ieee).
- Lu Y, Musalem A, Olivares M, Schilkret A (2013) Measuring the effect of queues on customer purchases. *Management Science* 59(8):1743–1763.
- Lucas RE (1976) Econometric policy evaluation: A critique. *Carnegie-Rochester conference series on public policy*, volume 1, 19–46 (Elsevier).
- Mani V, Kesavan S, Swaminathan JM (2015) Estimating the impact of understaffing on sales and profitability in retail stores. *Production and Operations Management* 24(2):201–218.
- Marinesi S, Girotra K (2013) Information acquisition through customer voting systems. *Working Paper*.
- Mihm J, Schlapp J (2015) Sourcing innovation: Public and private feedback in contests. *Working Paper*.
- Milgrom PR, Weber RJ (1985) Distributional strategies for games with incomplete information. *Mathematics of Operations Research* 10(4):619–632.
- Moldovanu B, Sela A (2001) The optimal allocation of prizes in contests. *American Economic Review* 542–558.
- Moon K, Bimpikis K, Mendelson H (2016) Randomized markdowns and online monitoring. *Management Science, Forthcoming*.
- Moreno A, Terwiesch C (2014) Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research* 25(4):865–886.
- Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Science* 56(7):1180–1197.
- Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science* 54(1):41–55.

- Olivares M, Weintraub GY, Epstein R, Yung D (2012) Combinatorial auctions for procurement: An empirical study of the chilean school meals auction. *Management Science* 58(8):1458–1481.
- Rust J (1987) Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.
- Schotter A, Weigelt K (1992) Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results. *The Quarterly Journal of Economics* 511–539.
- Taylor CR (1995) Digging for golden carrots: An analysis of research tournaments. *The American Economic Review* 872–890.
- Terwiesch C, Loch CH (2004) Collaborative prototyping and the pricing of custom-designed products. *Management Science* 50(2):145–158.
- Terwiesch C, Ulrich KT (2009) *Innovation tournaments: Creating and selecting exceptional opportunities* (Harvard Business Press).
- Terwiesch C, Xu Y (2008) Innovation contests, open innovation, and multiagent problem solving. *Management science* 54(9):1529–1543.
- Wooten JO, Ulrich KT (2016a) Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments. *Production and Operations Management, Forthcoming*.
- Wooten JO, Ulrich KT (2016b) The impact of visibility in innovation tournaments: Evidence from field experiments. *Production and Operations Management, Forthcoming*.
- Xu Y, Armony M, Ghose A (2016) The effect of online reviews on physician demand: A structural model of patient choice. *Working Paper*.
- Yang Y, Chen PY, Pavlou P (2009) Open innovation: An empirical study of online contests. *ICIS 2009 Proceedings* 13.
- Yu Q, Allon G, Bassamboo A (2016) How do delay announcements shape customer behavior? An empirical study. *Management Science, Articles in Advance*.

# Appendices

## A. Image Comparison Algorithm – SIFT

SIFT is an algorithm used to detect and describe local features in images proposed by Lowe (1999).

Broadly speaking, the algorithm consists of four steps:

- 1. Extracting SIFT Feature Vectors:** From a pair of design images (A and B), we extract descriptors of the key points by identifying *SIFT feature vectors* in scale space, which robustly capture the structural properties of the images<sup>24</sup>.
- 2. Matching SIFT Feature Vectors:** For each feature vector  $D_i^A$  in image A, we calculate its shortest Euclidean distance ( $d(D_i^A, D_j^B)$ ) to each of the SIFT feature vector in image B ( $D_j^B$ ). Features  $D_i^A$  and  $D_k^B$  are defined as a matched pair if and only if the ratio  $\frac{d(D_i^A, D_k^B)}{d(D_i^A, D_j^B; \forall j)}$  is always less than  $2/3$ <sup>25</sup>.
- 3. Computing the Similarity Ratio:** After obtaining the number of matched feature-vector pairs, we can calculate the percentage of matched SIFT features relative to the total number of SIFT features in images A and B as  $\gamma_{A,B} = \frac{N_{A,B}^m}{\min\{N_A, N_B\}}$ , where  $N_{A,B}^m$  is the number of matched pairs between image A and image B, and  $N_A$  ( $N_B$ ) is the total number of feature vectors extracted from image A (B).
- 4. Classifying the Image Pair as Similar or Different:** Finally, we classify the image pair (image A and image B) as either similar or different based on the Similarity Ratio,  $\gamma_{A,B}$ . The higher the ratio is, the more similar the two images are. If the ratio is 1, the two images are exactly the same. In our empirical analysis, we classify a pair of submissions as similar if  $\gamma_{A,B} \geq 0.4$ .

Figure A.1 provides an example of the computed Similarity Ratios among six designs. As we can see, the last three designs are relatively similar to each other, and they have higher pairwise similarity ratios (0.473, 0.522, 0.789 respectively).

<sup>24</sup> The scale-invariant features are efficiently identified by a staged filtering approach. In the first stage, the algorithm identifies key locations in scale space by looking for locations that are maxima or minima of a difference-of-Gaussian function. Then, each point is used to generate a feature vector that describes the local image region sampled relative to its scale-space coordinate frame.

<sup>25</sup> This is the default threshold in Lowe’s paper (1999). The approach described here is essentially the nearest-neighbor approach.



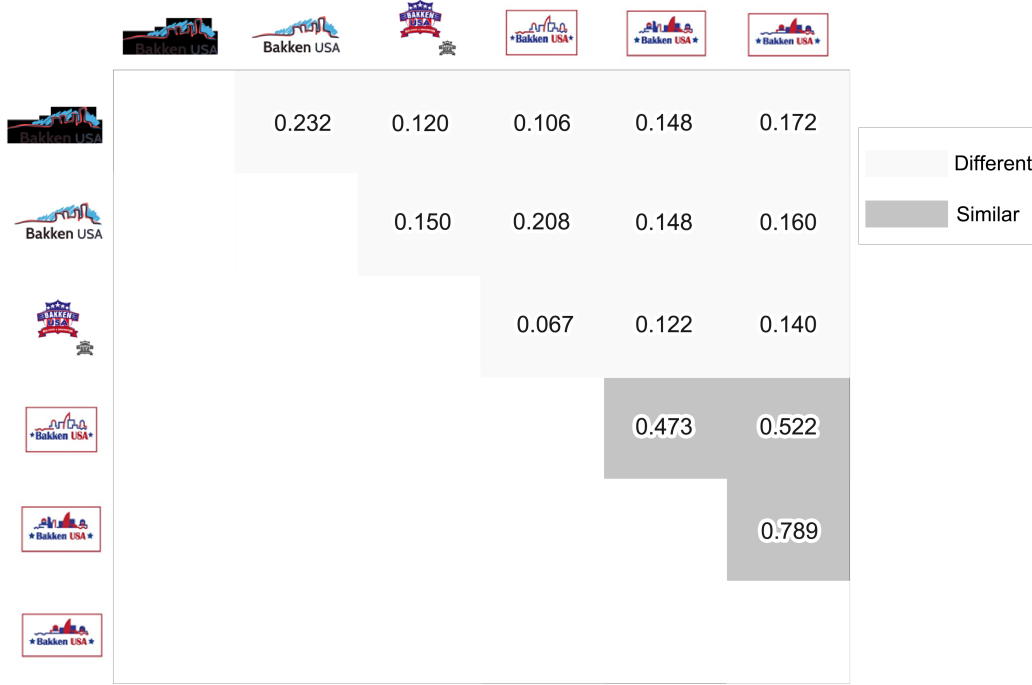


Figure A.1 Similarity Ratio Matrix for Six Designs

## B. Derivation of the MPE for the Structural Model (Equation 9 and 10)

In the proposed structural model, a strategy  $\sigma^* = \{\rho^*, \lambda^*\}$  representing a MPE equilibrium is characterized by, at any  $t$ , for any incumbent  $i$ :

$$\rho_{i,t}^* = \arg \max_{a_{i,t}} \left\{ U_{i,t}(a_{i,t}, \epsilon_{i,t}) + \beta \sum_{x_{i,t+1}, s_{t+1}} \left[ \int V_{i,t+1}^{\sigma^*}(x_{i,t+1}, s_{t+1}, \epsilon_{i,t+1}) p_{\epsilon} d(\epsilon_{i,t+1}) \right] p^{\sigma^*}(x_{i,t+1}, s_{t+1} | x_{i,t}, s_t, a_{i,t}) \right\},$$

and for any potential entrant  $j$ :

$$\lambda_{j,t}^* = \arg \max_{d_{j,t}} \left\{ U_{j,t}^e(d_{j,t}, \epsilon_{j,t}) + I(d_{j,t} = 1) \cdot \beta \sum_{x_{j,t+1}, s_{t+1}} \left[ \int V_{j,t+1}^{\sigma^*}(x_{j,t+1}, s_{t+1}, \epsilon_{j,t+1}) p_{\epsilon} d(\epsilon_{j,t+1}) \right] p^{e,\sigma^*}(x_{j,t+1}, s_{t+1} | s_t) \right\}.$$

where  $p^{\sigma}(x_{i,t+1}, s_{t+1} | x_{i,t}, s_t, a_{i,t}) = p(x_{i,t+1} | x_{i,t}, a_{i,t}) \cdot p^{\sigma}(s_{t+1} | s_t, x_{i,t}, a_{i,t})$  and  $p^{e,\sigma}(x_{j,t+1}, s_{t+1} | s_t) = p^e(x_{j,t+1}) * p^{\sigma}(s_{t+1} | s_t)$ , in which  $p^{\sigma}(s_{t+1} | s_t, x_{i,t}, a_{i,t}) = \sum_{a_t, d_t} p(s_{t+1} | s_t, a_t, d_t) \cdot Pr(a_t, d_t | s_t, \sigma_t)$ .

Representing MPE in probability space<sup>26</sup> (Milgrom and Weber 1985), we get equilibrium CCP  $\mathbf{P}^*$  as a fixed point  $\mathbf{P}^* = \mathbf{\Gamma}(\mathbf{P}^*)$ , where the function  $\mathbf{\Gamma}$  is the creators' *best response probability function*, with the  $t^{\text{th}}$  element being  $\mathbf{\Gamma}_t(\mathbf{P}) = \{\Gamma_t(\mathbf{P}), \Gamma_t^e(\mathbf{P})\}$  (with and without superscripts "e" denoting potential entrants' entry choices and incumbents' follow-up choices respectively). For incumbents,  $\mathbf{\Gamma}_t = \{\Gamma_{i,t}(\mathbf{P})\}_{i \in N_t}$ , where  $\Gamma_{i,t}(\mathbf{P}) = \{\Gamma_{i,t}(a_{i,t}|x_{i,t}, s_t; \mathbf{P}_{-i})|a_{i,t} \in \mathcal{A}\}$ , in which,

$$\Gamma_{i,t}(a_{i,t}|x_{i,t}, s_t; \mathbf{P}_{-i}) = \int I(a_{i,t} = \arg \max_{a_{i,t}} \{v_{i,t}^{\mathbf{P}}(x_{i,t}, s_t, a_{i,t}) + \epsilon_{i,t}(a_{i,t})\}) p_\epsilon(\epsilon_{i,t}) d\epsilon_{i,t}. \quad (20)$$

In the equation above,  $v_{i,t}^{\mathbf{P}}(x_{i,t}, s_t, a_{i,t}) = -c(a_{i,t}) + \beta \sum_{x_{i,t+1}, s_{t+1}} \left[ \int V_{i,t+1}^{\mathbf{P}}(x_{i,t+1}, s_{t+1}, \epsilon_{i,t+1}) p_\epsilon d(\epsilon_{i,t+1}) \right] p^\sigma(x_{i,t+1}, s_{t+1}|x_{i,t}, s_t, a_{i,t})$  is incumbents' *choice specific value function*. Similarly, for potential entrants,  $\mathbf{\Gamma}_t^e = \{\Gamma_{j,t}^e(\mathbf{P})\}_{j \in M_t}$ , where  $\Gamma_{j,t}^e(\mathbf{P}) = \{\Gamma_{j,t}^e(d_{j,t}|s_t; \mathbf{P}_{-j})|d_{j,t} \in \mathcal{D}\}$ , in which,

$$\Gamma_{j,t}^e(d_{j,t}|s_t; \mathbf{P}_{-j}) = \int I(d_{j,t} = \arg \max_{d_{j,t}} \{v_{j,t}^{e,\mathbf{P}}(s_t, d_{j,t}) + \epsilon_{j,t}(d_{j,t})\}) p_\epsilon(\epsilon_{j,t}) d\epsilon_{j,t}, \quad (21)$$

where  $v_{j,t}^{e,\mathbf{P}}(s_t, d_{j,t})$  is the potential entrants' choice specific value function.

Assuming both the incumbents' private shock ( $\epsilon_{i,t}$ ) and potential entrants' private shock ( $\epsilon_{j,t}$ ) follow the Type I extreme value distribution (Rust 1987), we can get a closed-form solution for the equilibrium CCP ( $\mathbf{P}^*$ ), i.e., incumbent  $i$  follows Equation 9, and potential entrant  $j$  follows Equation 10.

### C. Keane and Wolpin's Simulation and Interpolation Method

To deal with the "curse of dimensionality", we adopt Keane and Wolpin's simulation and interpolation method to reduce the computational burden (Keane and Wolpin 1994). For each period  $t$  starting from the last decision period  $T$ , we first sample a subset of frequently visited state points  $(x_{i,t}, s_t)$ . For each of these sampled state points, we solve the MPE backwards and calculate the value function exactly, given the next period's value functions. That is, we calculate the third-period exact value functions "exactly" from creator state-specific utility in the terminal period and the state transition probabilities; the second-period exact value functions are then calculated "conditionally exactly" from both the "exact" and interpolated third-period value functions and

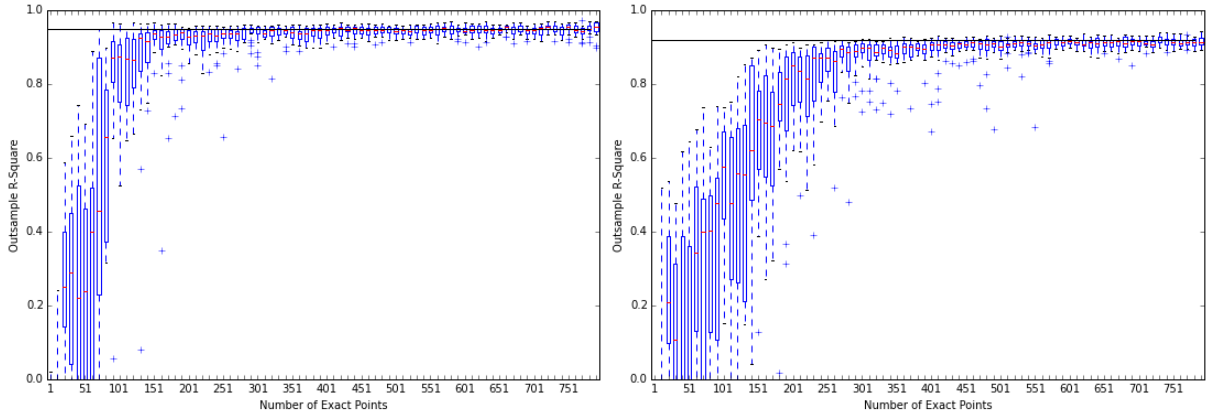
<sup>26</sup> Associated with  $\sigma_t$  we can define a set of CCP for incumbents and potential entrants in period  $t$ ,  $P_t^\sigma = \{\{P_{i,t}^\sigma(a_{i,t}|x_{i,t}, s_t)\}, \{P_{j,t}^\sigma(d_{j,t}|s_t)\}|i \in N_t, j \in M_t\}$  such that,

$$P_{i,t}^\sigma(a_{i,t}|x_{i,t}, s_t) = \int I\{\rho_{i,t}(x_{i,t}, s_t, \epsilon_{i,t}) = a_{i,t}\} p_\epsilon(\epsilon_{i,t}) d\epsilon_{i,t}, \text{ and}$$

$$P_{j,t}^\sigma(d_{j,t}|s_t) = \int I\{\lambda_{j,t}(s_t, \epsilon_{j,t}) = d_{j,t}\} p_\epsilon(\epsilon_{j,t}) d\epsilon_{j,t}.$$

Note that the probabilities  $P_{i,t}^\sigma(a_{i,t}|x_{i,t}, s_t)$  and  $P_{j,t}^\sigma(d_{j,t}|s_t)$  represent the expected behavior of a creator from the point of view of the rest of the creators when this creator follows strategy  $\sigma_t$ .

the state transition probabilities. Using these “exact value functions”, we interpolate the value functions at other state points. The so computed value functions in the current period  $t$  are then used in the calculation of the exact value functions in period  $t - 1$ , and so on and so forth. This interpolation step provides a good approximation for the value functions: As we can see from Figure C.1, the out-sample R-squares can be as high as 95% (92%) for second (third) period’s value functions when the number of sampled exact points is sufficiently large. Note that we get a higher out-sample R-square in the second-period interpolation. This is because the number of second period state combinations is much smaller than that of the third period. That is, with the same number of sampled exact points, we are computing proportionally more exact points for the second period. Here, we sample 600 exact points for the interpolation.



(a)  $t=2$  (b)  $t=3$   
**Figure C.1 Out-sample R-square for Value Function Interpolation**

## D. Frequency Estimates for Action-Specific Transitions

Frequency estimates for the rating distribution of new entrants' first submission(s) and the action-specific state transition probabilities among states up to 5-star are reported below.

**Table D.2 New Entrants Rating Distribution**

NA	1-Star	2-Star	3-Star	4-Star	5-Star
0.23	0.109	0.172	0.265	0.172	0.052

**Table D.3 Rating Improvement Resulting from "Revise" Action**

		Post Rating					
		NA	1-Star	2-Star	3-Star	4-Star	5-Star
Base Rating	NA	0.869	0.035	0.026	0.044	0.026	0
	1-Star	0	0.615	0.187	0.121	0.066	0.011
	2-Star	0	0	0.659	0.239	0.09	0.012
	3-Star	0	0	0	0.785	0.177	0.038
	4-Star	0	0	0	0	0.861	0.139
	5-Star	0	0	0	0	0	1

**Table D.4 Rating Improvement Resulting from "Redesign" Action**

		Post Rating					
		NA	1-Star	2-Star	3-Star	4-Star	5-Star
Base Rating	NA	0.771	0.032	0.039	0.087	0.055	0.016
	1-Star	0	0.639	0.132	0.118	0.083	0.028
	2-Star	0	0	0.649	0.235	0.095	0.021
	3-Star	0	0	0	0.783	0.158	0.059
	4-Star	0	0	0	0	0.869	0.131
	5-Star	0	0	0	0	0	1

**Table D.5 Rating Improvement Resulting from "Do-both" Action**

		Post Rating					
		NA	1-Star	2-Star	3-Star	4-Star	5-Star
Base Rating	NA	0.679	0.038	0.094	0.094	0.038	0.057
	1-Star	0	0.551	0.192	0.167	0.077	0.013
	2-Star	0	0	0.523	0.189	0.212	0.076
	3-Star	0	0	0	0.607	0.301	0.092
	4-Star	0	0	0	0	0.783	0.217
	5-Star	0	0	0	0	0	1

## E. Regression Results for Robustness Checks

**Table E.6 The Percentage of High Type Among New Entrants With Previous Feedback**

Dependent variable: $\% \Delta H_{qt}$			
$(No. 1-Star)_{qt-1}$	-0.001	(0.001)	
$(No. 2-Star)_{qt-1}$	-0.0003	(0.001)	
$(No. 3-Star)_{qt-1}$	0.0002	(0.001)	
$(No. 4-Star)_{qt-1}$	-0.001	(0.001)	
$(No. 5-Star)_{qt-1}$	0.001	(0.002)	
$(No. Submissions)_{qt-1}$	0.001*	(0.0005)	
$(No. Creators)_{qt-1}$	-0.002	(0.001)	
Time Dummies			Yes
Contest-level Fixed Effect			Yes
Observations	5,607	R <sup>2</sup>	0.002
Adjusted R <sup>2</sup>	0.002	F Statistic	1.244 (df = 7; 4793)

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001; the numbers in parenthesis are standard errors.

**Table E.7 Incumbent Follow-up Actions Multinomial Logit Model and Heterogeneity Evidence**

	Depend Variable: $Action_{iqt}$					
	<i>Re-design</i>		<i>revision</i>		<i>do-both</i>	
<b>Heterogeneity Variables</b>						
$I(H)_i$	-0.162*	(0.076)	-0.043	(0.048)	-0.008	(0.075)
$\%H_{qt-1}$	0.322	(0.307)	0.045	(0.188)	-0.092	(0.291)
<b>Individual-Level Variables</b>						
$(No. Submissions)_{iqt-1}$	0.011	(0.014)	0.080***	(0.006)	0.075***	(0.009)
$AvgRating_{iqt-1}$	0.121	(0.135)	0.108	(0.075)	0.140	(0.120)
$BestRating_{iqt-1}$	0.271***	(0.080)	0.352***	(0.048)	0.189*	(0.076)
$SecondBestRating_{iqt-1}$	-0.101	(0.088)	-0.034	(0.051)	-0.077	(0.081)
<b>Contest-Level Variables</b>						
$Award_q(\$)$	0.001	(0.0004)	0.001***	(0.0002)	0.001***	(0.0003)
$(No. 1-Star)_{qt-1}$	0.010***	(0.003)	0.010***	(0.002)	0.011***	(0.003)
$(No. 2-Star)_{qt-1}$	0.009**	(0.003)	0.008***	(0.002)	0.011***	(0.003)
$(No. 3-Star)_{qt-1}$	0.005	(0.003)	0.008***	(0.002)	0.006*	(0.003)
$(No. 4-Star)_{qt-1}$	0.002	(0.004)	0.005*	(0.003)	0.014***	(0.004)
$(No. 5-Star)_{qt-1}$	-0.012	(0.009)	-0.027***	(0.005)	-0.030***	(0.008)
$(No. Submissions)_{qt-1}$	-0.007*	(0.003)	-0.002	(0.002)	-0.007*	(0.003)
$(No. Creators)_{qt-1}$	0.005	(0.006)	-0.018***	(0.004)	-0.010	(0.006)
Time Dummies						Yes
Observations	24,085		R <sup>2</sup>		0.041	
Log Likelihood	-14,771.670		LR Test		1,250.716***	(df = 60)

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001; the numbers in parenthesis are standard errors.