Forthcoming at *Manufacturing & Service Operations Management*
manuscript MSOM-22-344

Product Development in Crowdfunding: Theoretical and Empirical Analysis

Sıdıka Tunç Candoğan
UCL School of Management, University College London, London, UK, E14 5AA, sidika.tunc.16@ucl.ac.uk

Philipp B. Cornelius
Rotterdam School of Management, Erasmus University Rotterdam, 3062 PA Rotterdam, Netherlands, cornelius@rsm.nl

Bilal Gokpinar
UCL School of Management, University College London, London, UK, E14 5AA, b.gokpinar@ucl.ac.uk

Ersin Körpeoğlu
UCL School of Management, University College London, London, UK, E14 5AA, e.korpeoglu@ucl.ac.uk

Christopher S. Tang
Anderson School of Management, University of California Los Angeles, Los Angeles, CA, 90095, chris.tang@anderson.ucla.edu

Problem definition: Crowdfunding goes beyond raising funds. Entrepreneurs often use crowdfunding to solicit feedback from customers in order to improve their products, and may therefore prefer to launch their crowdfunding campaigns using basic versions of their products with fewer features. However, customers may not be persuaded by a campaign if the product appears to be underdeveloped. In view of this trade-off, a key question for entrepreneurs is how much to develop a product before launching a crowdfunding campaign. **Methodology/results:** Analyzing a game-theoretical model and testing its predictions empirically, we study: 1) how the development level of a product at campaign launch, measured by the initial number of product features, influences whether customers will make comments that help entrepreneurs improve the product; 2) whether entrepreneurs continue to improve the product during the campaign; and 3) whether the campaign is successful. We show that, as the number of product features at campaign launch increases, the likelihood that customers will make comments and that the product will be improved during the campaign first increases but then decreases. Furthermore, the likelihood of campaign success first increases but then decreases with the number of product features at campaign launch. Finally, by analyzing the interactions between customer feedback, product improvement, and campaign success, we show that customer feedback motivates entrepreneurs to improve the product during the campaign. Moreover, entrepreneurs should take account of the initial number of features and customer feedback when improving the product, because otherwise product improvements can harm campaign success. **Managerial implications:** Our study provides practical insights on how entrepreneurs can use crowdfunding to aid product development and improvement. Specifically, entrepreneurs should avoid overdeveloping their products before crowdfunding campaigns because, as well as decreasing the chance of campaign success, this could hinder their ability to save development costs (e.g., market research costs) through involving customers in product development.

Key words: product development, entrepreneurship, innovation, natural language processing

(a) Initial product. (b) Customer comment. (c) Product improvement.

Figure 1 Example of initial product, customer comment, and product improvement.

1. Introduction

The internet enables entrepreneurs to use crowdfunding to raise funds from a large number of people for projects ranging from those developed through social entrepreneurship to for-profit enterprises. Recent research and practice suggest that, as well as being an important financial instrument (e.g., Hu et al. 2015, Belavina et al. 2020, Chakraborty and Swinney 2021), crowdfunding can be used by entrepreneurs (hereafter *creators*) as a mechanism for involving customers in product development, enabling them to improve their products during their crowdfunding campaigns (e.g., Mollick 2016, Cornelius and Gokpinar 2020). Because customer involvement in product development can lead to significant cost savings (e.g., Thomke and Bell 2001, Loch and Kavadias 2008), creators may consider launching their crowdfunding campaign using a basic version of a product with fewer features so as to leave room for improvements. However, if a product appears to be underdeveloped, customers may be discouraged from contributing to the campaign and from commenting, making it less likely that creators will improve the product and that the campaign will be successful (i.e., reach the funding goal). Considering this trade-off, we study how the number of features a product has at campaign launch affects customer feedback, product improvement, and campaign success.

To understand how customers can contribute to product improvement during a crowdfunding campaign, we can consider the following example from Kickstarter—a global crowdfunding platform that has raised \$5 billion for entrepreneurs over the last decade (Kickstarter 2021f). In October 2015, the "onomo" team launched a campaign for an innovative bike navigation device, HAIZE (see Figure 1(a); Kickstarter 2021a). During the campaign, customers suggested that HAIZE might include a wristband so that they could also use the product when not on their bike (see Figure 1(b)). In response to these customer suggestions, the creators added a wristband to the product and revised the campaign description accordingly (see Figure 1(c)). In our interview with them, the creators of HAIZE explained this process as follows: "It's definitely very efficient for that kind of [market] research... the idea, for example, of adding a wristband to the device, it was always like floating... we're not sure if we should do this [or not]... But then, we began to receive very relevant testimonies of people who were having amazing ideas [about a wristband]..."

As this example aptly shows, creators can make a strategic choice to launch a crowdfunding campaign using a basic product with fewer features and can improve it during the campaign by adding new features.[1] To shed light on this process, we investigate the following research questions: How does the initial number of product features affect (Q1) customer feedback, (Q2) the likelihood that the product will be improved during the campaign, and (Q3) the success of the campaign?

To answer these questions, and inspired by the crowdfunding practice and literature (e.g., Hu et al. 2015, Belavina et al. 2020, Chakraborty and Swinney 2021), we first generate theoretical predictions by analyzing a parsimonious game-theoretical model of a "reward-based, all-or-nothing" crowdfunding campaign that takes account of customer feedback and the creator's product improvement decisions.[2] In such a campaign, creators solicit funds from customers to finance the launch of a product. To this end, creators announce the initial number of product features, a funding goal, and a pledge price. Having considered these, customers then decide whether to pledge money in return for the product, and after pledging, they can make comments to induce the creators to improve the product further (customers cannot comment before pledging; see Kickstarter 2021g). When the creators observe comments, they decide whether to improve the product. If improvements are made, other potential customers will observe the improved product before making their pledging decisions. By the end of the campaign, if the total amount pledged reaches the funding goal, the campaign is successful; the creators receive the funds raised and products are then produced and delivered to customers. If the total amount pledged fails to reach the funding goal, the campaign fails; the creators receive and deliver nothing, and the customers are fully refunded.

We empirically test the theoretical predictions from our model in a unique large-scale data set from Kickstarter. Our data set contains detailed information about campaign characteristics along with product descriptions from the beginning and end of each campaign. We create a measure of the number of product features using detailed product descriptions and using an unsupervised natural language processing technique—latent Dirichlet allocation (LDA; Blei et al. 2003).[3] We address possible endogeneity concerns in our empirical analysis by exploiting a policy change on Kickstarter which reduced the minimum eligibility requirements for launching a campaign and thereby provided an exogenous shock to the initial number of product features.

[1] In our main analysis, we define "improvement" as the process of the creators adding new features. It is possible that a product can also be improved by eliminating some features. We analyze this case in §EC.5 in the Online Appendix.

[2] There are also other forms of crowdfunding with respect to the type of reward and type of funding. We refer the reader to Chen et al. (2020) for a review of other forms of crowdfunding.

[3] LDA has been used by marketing and operations scholars for example to extract product features (Toubia et al. 2019) or generate a measure of a firm's innovation (Bellstam et al. 2020) from textual data.

We first analyze how the initial number of product features affects commenting by customers. We show that the likelihood of customers commenting first increases but then decreases with the initial number of features. This is because a product with more features increases the customers' utility and hence increases pledging and commenting probabilities up to a point. However, if there are too many features, customers can be overwhelmed and refrain from pledging and commenting.

We next analyze how the initial number of product features affects the likelihood of a product being improved with new features during the campaign. Because it is costly to make improvements, one might expect that the more features a product initially has, the less likely the creator will be to make any further improvement during the campaign. However, our analysis reveals an opposite effect: with more (but not too many) product features at campaign launch, customers are more likely to pledge and comment, and hence the creator is more likely to receive comments that can be used to improve the product. This increases the likelihood of product improvement during the campaign. However, above a certain initial number of features, the likelihood of product improvement decreases because of the additional cost of the new features.[4]

To address our third research question, we analyze the impact of the initial number of features on campaign success. We show that the likelihood of campaign success first increases but then decreases with the initial number of product features. This is because having more features increases the value of the product and hence increases the customer's utility up to a certain point. However, they then reduce the customer's utility because the product becomes too complex.

We show the robustness of our empirical results by considering the possible interplay between customer feedback, product improvement, and campaign success. This analysis also enables us to generate additional results. We show that customer feedback has a positive impact on the likelihood of product improvement and on the likelihood of campaign success. Furthermore, our findings suggest that creators should take account of the initial number of features and customer feedback when improving the product. If they do not, then unsolicited, creator-driven product improvements may harm the chance of campaign success.

Related Literature. As a phenomenon that has emerged quickly, crowdfunding has caught the attention of entrepreneurs, managers, and business scholars. The literature on crowdfunding is therefore relatively new but growing. We discuss theoretical and empirical studies of crowdfunding before summarizing our contributions to this literature.

In the crowdfunding literature, Hu et al. (2015) study reward-based, all-or-nothing crowdfunding to analyze whether creators should offer a single reward or multiple rewards. Follow-up studies

[4] For instance, while responding to a customer comment asking for an extra USB port for a 360° camera, one creator on Kickstarter explained why this improvement was not feasible by saying: "Yes, I would love to have USB3, or USB type C... What you might not know fully is that it is a serious additional cost..." (Kickstarter 2021e).

(e.g., Du et al. 2017, Chakraborty and Swinney 2019, 2021, Burtch et al. 2020, Li et al. 2020) analyze the creators' other design decisions, including the funding goal, the pledge price, limited rewards, and the timing of referrals and contingent stimulus policies (e.g., limited-time offer). There are also theoretical studies that analyze how crowdfunding platforms can prevent misconduct (Strausz 2017, Belavina et al. 2020).[5] Empirical studies focus mainly on factors that influence the pledging decisions of customers and campaign success such as altruism, geographic proximity to creators, and creators' pre-campaign information sharing (e.g., Burtch et al. 2013, Mollick 2014, Agrawal et al. 2015, Lin and Viswanathan 2016, Kuppuswamy and Bayus 2017, Wei et al. 2020). Other empirical papers study broader aspects of crowdfunding such as the similarity between evaluations made by crowdfunding customers and experts (Mollick and Nanda 2016), the impact of crowdfunding on creators' ability to reach venture capital investors (Sorenson et al. 2016), and differences between the pledge price and the post-campaign retail price (Blaseg et al. 2020). Cornelius and Gokpinar (2020) show that crowdfunding campaigns are more likely to be successful if they have greater customer involvement and that the chance of campaign success increases with product improvements that are driven by customer feedback. For detailed reviews of this literature, we refer the reader to Allon and Babich (2020) and Chen et al. (2020).

While existing research has significantly improved our understanding of crowdfunding as a new form of financing (e.g., Hu et al. 2015, Belavina et al. 2020) and a customer interaction mechanism (e.g., Mollick 2016, Cornelius and Gokpinar 2020), the product development decisions made by creators in crowdfunding have not received empirical or theoretical attention in the literature. Our paper fills this gap by providing a nuanced understanding of crowdfunding as a product development mechanism. Specifically, inspired by practice and theoretical models in the crowdfunding literature, we first construct a theoretical model that includes customers' commenting and creators' product improvement decisions during campaigns, and then we test this theory empirically. We contribute to the literature by showing that creators should avoid overdeveloping their products before campaigns because this hinders both the chance of campaign success and the opportunity for the creators to save from development costs (e.g., market research costs) by involving customers in product development. We further show that improvised product improvements that are not based on customer feedback can harm the chance of campaign success. To our knowledge, our study is the first in the crowdfunding literature to combine theoretical and empirical analyses.

[5] There are other theoretical studies that ask broader questions about crowdfunding such as when to use different forms of crowdfunding (e.g., Belleflamme et al. 2014, Bi et al. 2019) or how crowdfunding interacts with traditional financing sources (e.g., Roma et al. 2018, Babich et al. 2021). Recently, Chemla and Tinn (2020) analyze the value of crowdfunding in testing the potential market. For a detailed review, we refer the reader to Chen et al. (2020).

Our paper also relates to the new product development (NPD) and the entrepreneurship literature on when to launch a product by considering different trade-offs.[6] Specifically, assuming that a more developed product always increases the customer's utility, the majority of studies in this literature (e.g., Cohen et al. 1996, Özer and Uncu 2013, Gao et al. 2021) investigate the extent to which a product launch should be delayed by considering the risk of losing the first-mover advantage. Similarly, the entrepreneurship literature (e.g, Choi et al. 2008, Lévesque et al. 2009) often focuses on the trade-off between learning and increased competition. Bhaskaran et al. (2020) consider a different trade-off where launching a basic and immediately available version of a product brings earlier revenues (as opposed to delaying the launch for a more developed version), but this can negatively affect the perception of future versions of the product. Yoo et al. (2021) consider another trade-off where launching a test product with higher quality increases each customer's utility and hence sales, whereas launching a test product with lower quality increases the entrepreneur's opportunity to learn about customer preferences from sales (or no sales).

Unlike these papers, we analyze a setting where launching a crowdfunding campaign for a basic version of the product brings no early revenue or first-mover advantage. Nor does it facilitate sales-based learning about customer preferences. Therefore, we identify a novel product development trade-off which is unique to crowdfunding and contributes to the NPD literature in several ways. First, we show that adding too many features to the product before a crowdfunding campaign can reduce the likelihood of campaign success, which implies that the customer's utility does not always increase with how developed the product is. Thus, the common assumption in the literature about product launches does not seem to hold in crowdfunding settings. Second, we consider a setting where customers receive the final version of the product despite having pledged before the creator improved the product during the campaign. This unique feature of crowdfunding eliminates the risk of losing future value, unlike settings where customers stick with the product that they purchased even though the product is later improved. Despite this advantage, there are additional challenges for the creator because the creator receives pledges only when the campaign is successful, and improvements hinge on customers' voluntary feedback. Therefore, unlike the settings analyzed in the NPD literature, we show that, in crowdfunding, launching a product with more features solicits more feedback. In summary, our study expands the NPD literature on product launch decisions by focusing the novel product development and financing setting.

[6] Our study also relates to the broader NPD literature (e.g., Thomke and Bell 2001, Loch et al. 2001, Erat and Kavadias 2008, Sommer et al. 2009) which mainly focuses on operational decisions related to experimentation and testing (e.g., whether to test sequentially or in parallel) mostly to resolve technical uncertainty *before* product launch. This literature suggests that the cost of making changes and redesigning increases over time in a product development process. We build our theoretical model based on this result, and study an innovative setting where product development continues based on customer feedback *after* product launch (i.e., launch of a crowdfunding campaign).

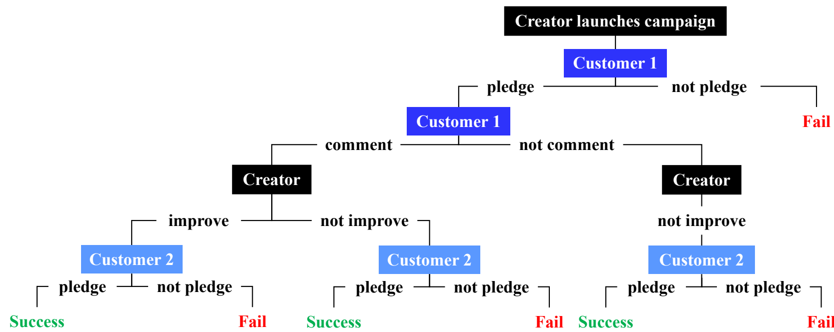


Figure 2 The sequence of decisions and events in a crowdfunding campaign where the product may be improved during the campaign.

2. Theoretical Model and Analysis

We consider a reward-based crowdfunding campaign in which a creator elicits funds from customers to finance the launch of a new product, and each customer pledges to receive the product as a reward. We focus on an all-or-nothing setting where the creator sets a funding goal and where if the total amount pledged exceeds the funding goal by the end of the campaign, then the campaign is successful. In that case, the creator collects the money pledged and delivers products to those who made pledges. If the total amount pledged does not meet the funding goal, the campaign fails. In this case, the creator does not receive any funds and does not deliver any products, and the customers are fully refunded (e.g., Hu et al. 2015, Belavina et al. 2020).

As crowdfunding is a nascent research area and our aim is to use our theoretical results to develop testable hypotheses, we develop a parsimonious model. We build on the model of Hu et al. (2015) by incorporating the customer's commenting decision and the creator's product improvement decision, whereby we construct a four-stage game-theoretical model that involves one creator and two customers, as illustrated in Figure 2. We describe our model according to the sequence of events.

Stage 0: The creator launches a crowdfunding campaign. The creator launches the campaign with a funding goal G (>0) and a set of product features, where q_i (>0) represents the number of product features.[7] (Throughout the paper, subscript "i" stands for "initial.") We assume that the creator incurs an investment cost of $C_i \cdot q_i$, where $C_i \geq 0$, to launch a campaign for a product with q_i features (e.g., Chakraborty and Swinney 2021). The initial investment cost C_i corresponds to the cost of adding each feature to the product before the campaign (e.g., the cost of market research or concept generation for each feature). The creator sets the pledge price $p = G/2$ so that the campaign is successful if and only if both customers make pledges (Hu et al. 2015).[8]

[7] We take the initial number of features q_i as an exogenous variable in our theoretical analysis to be consistent with our empirical analysis where we investigate the impact of the initial number of features on various metrics.

[8] As a supplementary analysis, we also consider a model where two customers (instead of just customer 2) arrive after customer 1 but where the funding goal G is still $2p$. In this case, it is still possible that the campaign will be successful

Each customer has a base valuation (hereafter, "valuation") v_j, $j = \{1, 2\}$, which measures customer j's marginal willingness to pay for the product, where v_j's are independent across customers (e.g., Belavina et al. 2020) and drawn from a Uniform distribution with parameters 0 and 1 (e.g., Krishnan and Ramachandran 2011, Belleflamme et al. 2014). The value a customer receives from the product depends both on the customer's valuation v_j and on the final number of product features q_f. (Throughout the paper, subscript "f" stands for "final."). Specifically, q_f has a mixed effect on the customer's value from the product. On the one hand, a higher q_f means that customers can find more things they like and assign a higher value to the product (e.g., Brown and Carpenter 2000). On the other hand, a higher q_f increases the complexity of the product due to it having a greater number of features as well as more interactions between these features. Such complexity can overwhelm customers (e.g., Mick and Fournier 1998) or can lead to customer anxiety (e.g., Castaño et al. 2008, Goodman and Irmak 2013), and can therefore reduce the value a customer receives from the product (e.g., Thompson et al. 2005). Combining these two opposing effects, we assume that a customer's effective value from the product is $v_j \cdot (q_f - b \cdot q_f^2)$, where $b > 0$.[9]

Stage 1: Customer 1's pledging decision. In Stage 1, customer 1 with valuation v_1 arrives at the campaign. In addition to observing the pledge price p and the initial number of features q_i of the product from its detailed description, customer 1 anticipates the final number of features q_f of the product. Given this information, customer 1 decides whether to pledge or not, by comparing the customer's effective valuation $v_1 \cdot (q_f - b \cdot q_f^2)$ and the pledge price p.

If customer 1 does not pledge in Stage 1, the campaign fails, and both customers receive a reservation value of 0. If customer 1 pledges, then the next stage commences.

Stage 2: Customer 1's commenting decision. Customer 1 decides whether to make comments, and the number of comments made may depend on the initial number of features q_i.[10] Specifically, the number of comments may increase with the initial number of features q_i because more features can stimulate more ideas, or the number of comments may decrease with q_i as more features can restrict the improvement potential of the product. To capture the potential relationship between q_i and the number of comments without imposing any directional effect, we assume that the number of comments that customer 1 makes is $N = q_i^n$, where $n \in \mathbb{R}$. This functional form enables us to

if two out of three customers pledge. Hence, no single customer is pivotal in determining whether the campaign will be successful. Our supplementary analysis of this case yields similar qualitative results to our main results.

[9] Here, we assume that any additional feature increases the value that each customer assigns to the product. It is possible that a customer may value some features but not others. In §EC.2 in the Online Appendix, we extend our main results to the case where each customer values a random fraction of q_i features.

[10] For ease of illustration, we assume that the customer's cost of commenting is negligible compared to the utility the customer can obtain from a potential improvement in the product. In §EC.3 in the Online Appendix, we extend our analysis to the case where the customer incurs some non-negligible cost when making comments.

capture situations where the number of comments N is increasing, constant, and decreasing in q_i when $n > 0$, $n = 0$, and $n < 0$, respectively.

Stage 3: The creator's product improvement decision. If customer 1 pledges in Stage 1 and makes N comments in Stage 2, then the creator decides whether or not to improve the product by adding q_u new features by considering the additional investment cost.[11] (Throughout the paper, subscript "u" stands for "upgrade", i.e., improvement.) When more comments are made, the creator may require less research about q_u or concept generation can be easier for the creator because of the higher level of information provided through the customer's comments. Therefore, we assume that the additional investment cost of q_u is decreasing in the number of comments N, and takes the form $\frac{C_u \cdot q_u}{N+1}$ ($C_u \geq 0$) for ease of illustration. (Note that our results continue to hold when the investment cost of improvement is independent of the number of comments.)

Stage 4: Customer 2's pledging decision. Customer 2 with valuation v_2 arrives at the campaign, observes the final number of features q_f along with the pledge price p, and decides whether to pledge or not. If customer 2 pledges along with customer 1, then the campaign is successful, and the creator receives pledges and delivers products to customers by incurring a per-unit production cost of $c \cdot q_f^2$ ($c > 0$). (Our results are qualitatively similar when the per-unit production cost is $c \cdot q_f$.) This production cost might, for instance, represent the cost of materials and labor (e.g., Guo and Zhang 2012, Hu et al. 2015). If customer 2 does not pledge, then the campaign fails.

2.1. Analysis of Sub-game Perfect Equilibrium

For any given pledge price p and initial number of features q_i, we determine the sub-game perfect equilibrium via backward induction.

First, in Stage 4, customer 2 with valuation $v_2 \sim \text{Uniform}(0,1)$ pledges if and only if customer 2's expected utility $U_2 = v_2(q_f - b \cdot q_f^2) - p \geq 0$ and observes that customer 1 has pledged in Stage 1, anticipating that the campaign can be successful only when both customers pledge.

ASSUMPTION 1. *$q_f - bq_f^2 > p$ for any q_f such that there exists a valuation $v_2 \in (0,1)$ which makes customer 2's expected utility U_2 positive.*

Under Assumption 1, if the creator improved the product in Stage 3, the final number of features $q_f = q_i + q_u$ and hence customer 2 pledges with probability $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$; otherwise, $q_f = q_i$ and hence customer 2 pledges with probability $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$.[12]

[11] While we take q_u as positive and exogenous in our main analysis, in §EC.5 in the Online Appendix, we extend our analysis to the case where q_u is endogenously determined and can be negative.

[12] Here, we assume that the impact of q_u on the value of the product equals to the impact of q_i. In §EC.4 in the Online Appendix, we extend our main results to the case where q_u can have a larger impact on the product value.

In Stage 3, if customer 1 pledged and made comments in Stages 1 and 2, the creator decides whether to improve the product during the campaign by comparing the creator's expected profit Π^I with improvement and expected profit Π^{NI} with no improvement. (Superscripts I and NI stand for "with improvement" and "with no improvement," respectively.) By taking account of customer 2's pledging probability $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$ and the per-unit production cost $c(q_i+q_u)^2$, the creator's expected profit with improvement is $\Pi^I = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(2p-2c(q_i+q_u)^2) - C_i q_i - \frac{C_u q_u}{N+1}$. Similarly, the creator's expected profit without improvement is $\Pi^{NI} = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(2p-2cq_i^2) - C_i q_i$. Thus, the creator improves the product during the campaign if and only if $\Pi^I \geq \Pi^{NI}$, i.e.,

$$I \equiv \frac{2p^2(1-b(2q_i+q_u))}{q_i(q_i+q_u)(1-b(q_i+q_u))(1-bq_i)} - 2c\left(2q_i+q_u - \frac{p}{(1-b(q_i+q_u))(1-bq_i)}\right) - \frac{C_u}{N+1} \geq 0. \quad (1)$$

In Stage 2, if customer 1 pledged in Stage 1, then this customer decides whether or not to make comments. First, suppose that condition (1) is violated. Then, regardless of making comments, customer 1 anticipates that the creator will not improve the product in Stage 3 and that customer 2 will pledge in Stage 4 with probability $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$. Thus, regardless of making comments or not, customer 1's expected utility is $U_1^C = U_1^{NC} = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(v_1(q_i-bq_i^2)-p)$. (Superscripts C and NC stand for "with commenting" and "with no commenting," respectively.) Because customer 1 is indifferent about making comments or not, both cases are in equilibria.

Now, suppose that condition (1) holds. Then, if customer 1 makes comments, this customer anticipates that the creator will improve the product in Stage 3 and customer 2 will pledge in Stage 4 with probability $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$. Thus, in this case, customer 1's expected utility is $U_1^C = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(v_1((q_i+q_u)-b(q_i+q_u)^2)-p)$. However, if customer 1 does not make comments, this customer anticipates that customer 2 will pledge in Stage 4 with probability $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$, and hence customer 1's expected utility is $U_1^{NC} = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(v_1(q_i-bq_i^2)-p)$. Noting that condition (1) can be satisfied only when $(q_i+q_u)-b(q_i+q_u)^2 > q_i - bq_i^2$, customer 1 makes comments in Stage 2 if and only if $U_1^C \geq U_1^{NC}$, i.e., $v_1 \geq \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)}$.

Finally, in Stage 1, customer 1's pledging decision depends on whether customer 1 anticipates an improvement in the product or not. First, suppose that condition (1) holds so that customer 1 anticipates an improvement and makes comments when $U_1^C \geq U_1^{NC}$. Then, customer 1 decides whether to pledge or not by comparing the expected utility U_1^P when pledging, where $U_1^P = U_1^C = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(v_1((q_i+q_u)-b(q_i+q_u)^2)-p)$, and her expected utility U_1^{NP} when not pledging, where $U_1^{NP}=0$. (Superscripts P and NP stand for "when pledging" and "when not pledging.") Thus, customer 1 pledges if and only if $U_1^P \geq U_1^{NP}$, i.e., $v_1 \geq \frac{p}{(q_i+q_u)-b(q_i+q_u)^2}$. Therefore, in a setting where customer 1 anticipates an improvement, customer 1 pledges *and* comments if

$$v_1 \geq \max\left\{\frac{p}{(q_i+q_u)-b(q_i+q_u)^2} \times \frac{p}{(q_i-bq_i^2)}, \frac{p}{(q_i+q_u)-b(q_i+q_u)^2}\right\} = \frac{p}{(q_i+q_u)-b(q_i+q_u)^2}.$$

This means that if condition (1) holds, customer 1 makes a comment when pledging.

Second, suppose that condition (1) is violated. Then, regardless of customer 1's commenting decision in Stage 3, customer 2 pledges in Stage 4 with probability $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$. In this case, customer 1's expected utility when pledging is $U_1^P=\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(v_1(q_i-bq_i^2)-p)$, and the expected utility when not pledging is $U_1^{NP}=0$. Thus, customer 1 pledges if and only if $v_1\geq\frac{p}{q_i-bq_i^2}$. Note that while making the decision, customer 1 considers customer 2's probability of pledging (i.e., $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$) because campaign success depends on the decisions of both customers.

By using the creator's and customers' rational strategies above, we next characterize the equilibrium outcomes. We aim to understand how the initial number of features affects customer feedback, product improvement, and campaign success. We first characterize the *ex-ante* probability that the product will be improved during the campaign, $\mathbb{P}(improve)$; and the *ex-ante* probability that the campaign will succeed, $\mathbb{P}(success)$. Consistent with these measures, we operationalize customer feedback by measuring the *ex-ante* probability that customer 1 will make comments, $\mathbb{P}(comment)$. Following the literature (e.g., Cornelius and Gokpinar 2020), we also characterize the expected number of comments $E[\#of comments]$. We present all proofs in §EC.1 in the Online Appendix.

Lemma 1. *(a) Suppose* $I\geq 0$. *Then,* $\mathbb{P}(comment)=\mathbb{P}(improve)=\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$, $E[\#of comments]=q_i^n\left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)$, *and* $\mathbb{P}(success)=\left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)^2$.
(b) Suppose $I<0$. *Then, either* $\mathbb{P}(comment)=\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$ *and* $E[\#of comments]=q_i^n\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)$ *or* $\mathbb{P}(comment)=E[\#of comments]=0$. *In both cases,* $\mathbb{P}(improve)=0$ *and* $\mathbb{P}(success)=\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)^2$.

Lemma 1(a) characterizes the case where condition (1) holds so that the creator is willing to improve the product during the campaign. In this case, customer 1 makes comments when pledging, and hence $\mathbb{P}(comment)=\mathbb{P}(improve)=\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$. Also, customer 2 pledges with probability $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$, and hence $\mathbb{P}(success)=\left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)^2$.

Lemma 1(b) shows that if condition (1) is violated, regardless of customer 1 making comments, $\mathbb{P}(improve)=0$ and $\mathbb{P}(success)=\left(\frac{q_i-bq_i-p}{q_i-bq_i^2}\right)^2$. As $q_f=q_i$, customer 1 is indifferent about making comments or not making comments, and hence we have two equilibria where either $\mathbb{P}(comment)=\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$ and $E[\#of comments]=q_i^n\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)$ or $\mathbb{P}(comment)=E[\#of comments]=0$. For ease of illustration, we focus on the equilibrium where customer 1 makes comments even if the customer anticipates that the product will not be improved, consistent with our observations on Kickstarter. This can be because the customer receives intrinsic benefit by making comments. It is also possible that there is some non-negligible cost of making comments, and hence the customer does not make any comments when anticipating no product improvement. In §EC.3 in the Online Appendix, we show the robustness of our theoretical predictions in this case.

2.2. Customer Feedback

We first answer our first research question (Q1) about customer feedback. The following proposition characterizes the impact of the initial number of features q_i on $\mathbb{P}(comment)$ and $E[\#of comments]$.

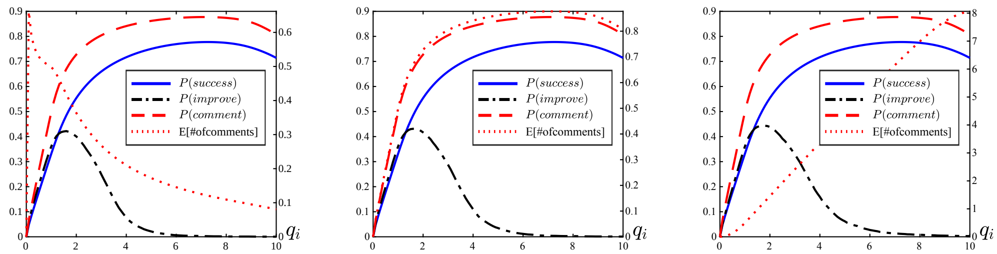
PROPOSITION 1. *Suppose that* $n \in \{n \in \mathbf{R} | \frac{\partial I}{\partial q_i} < 0$ *for any* $q_i\}$.[13] *There exists a threshold* $\overline{q_i}$ *on the initial number of features* q_i *such that:*
(a) When $q_i \leq \overline{q_i}$, $\mathbb{P}(comment)$ *is increasing in* $q_i \in (0, \frac{0.5}{b} - q_u)$ *and decreasing in* $q_i \in (\frac{0.5}{b} - q_u, \overline{q_i})$. *When* $q_i > \overline{q_i}$, $\mathbb{P}(comment)$ *is increasing in* $q_i \in (\overline{q_i}, \frac{0.5}{b})$ *and decreasing in* $q_i \in (\frac{0.5}{b}, \infty)$.
(b) There exist thresholds n' *and* n'' *such that (i) when* $q_i \leq \overline{q_i}$, $E[\#of comments]$ *is increasing in* q_i *if and only if* $n > n'$; *(ii) when* $q_i > \overline{q_i}$, $E[\#of comments]$ *is increasing in* q_i *if and only if* $n > n''$.

Proposition 1(a) shows that regardless of anticipating product improvement during the campaign or not (i.e., $q_i \leq \overline{q_i}$ or $q_i > \overline{q_i}$, respectively), the probability of customer 1 making comments increases with the initial number of features q_i as long as $bq_f < 0.5$. Otherwise, the probability of customer 1 making comments decreases with q_i. The intuition is that the customer's utility increases with the initial number of product features as long as the complexity of the product does not increase too much with the number of features.

Proposition 1(b) shows that given that $\mathbb{P}(comment) > 0$, the expected number of comments increases with q_i when n is large, and the expected number of comments decreases with q_i when n is small. The intuition is as follows. The expected number of comments depends on the probability of customer 1 making comments (i.e., $\mathbb{P}(comment)$) and the number of comments that customer 1 makes (i.e, q_i^n). As we show in Proposition 1(a), $\mathbb{P}(comment)$ first increases but then decreases with q_i; and q_i^n is increasing, constant or decreasing in q_i depending on the value of n. Therefore, when n is large, the increase in the number of comments dominates the possible decrease in the probability of commenting, and hence the expected number of comments increases with q_i. However, when n is small, the decrease in the number of comments dominates the possible increase in the probability of commenting, and hence the expected number of comments decreases with q_i.

Proposition 1 analyzes the impact of q_i on $\mathbb{P}(comment)$ and $E[\#of comments]$ for a given campaign, but our subsequent empirical analysis generates predictions about the average scenario over many campaigns. Thus, to better align our theoretical prediction with our empirical analysis, we conduct a numerical analysis to capture the theoretical predictions for the average scenario. Specifically, we calculate $\mathbb{P}(comment)$ and $E[\#of comments]$ for each q_i under 10,000 randomly generated instances according to the setting in Figure 3, and take the average. (Note that although

[13] This mild assumption is satisfied when n is non-negative or when n is negative but not too small. The precise condition on n is provided in the proof of Proposition 1. This assumption helps us avoid the unrealistic case where it is easier for the creator to improve the product further when q_i is larger.



(a) $n = -1$. (b) $n = 0$. (c) $n = 1$.

Figure 3 The impact of q_i on $E[\#of comments]$ (red dotted line), $\mathbb{P}(comment)$ (red dashed line), $\mathbb{P}(improve)$ (black dash-dotted line), and $\mathbb{P}(success)$ (blue solid line). The figure presents the average of 10,000 randomly generated instances where we select p from Uniform(0,1), c from Uniform(0,0.01), C_i, C_u, q_u and b from Uniform(0,0.1). The first y-axis represents the probabilities while the second y-axis represents the number of comments.

Proposition 1(a) shows that there can be double inverted U-shape relationship between q_i and $\mathbb{P}(comment)$, our numerical analysis shows that, on average, $\mathbb{P}(comment)$ first increases and then decrease with q_i.) Based on Proposition 1 and our numerical analysis, we establish the following hypothesis about $\mathbb{P}(comment)$ (hereafter H1).

Hypothesis 1. *As the initial number of product features increases, the likelihood that customers will make comments first increases and then decreases.*

Our theory suggests that the impact of the initial number of product features on $E[\#of comments]$ depends on how the initial number of product features affects the number of comments that a pledging customer makes (i.e., the parameter n). Because we do not impose any restrictive assumption on this parameter, the impact of the initial number of product features on $E[\#of comments]$ remains an empirical question that we answer in §3.

2.3. Probability of Product Improvement

Using Lemma 1, we next answer our second research question (Q2). The following proposition characterizes the impact of the initial number of features q_i on $\mathbb{P}(improve)$.

Proposition 2. *Let $\overline{q_i}$ be the threshold defined in Proposition 1. When $q_i \le \overline{q_i}$, $\mathbb{P}(improve)$ is increasing in $q_i \in (0, \frac{0.5}{b} - q_u)$ and decreasing in $q_i \in (\frac{0.5}{b} - q_u, \overline{q_i})$. When $q_i > \overline{q_i}$, $\mathbb{P}(improve) = 0$.*

One might expect that the higher the initial number of features q_i is, the less likely it is that the product will be improved during the campaign. However, Proposition 2 shows that when q_i is below a threshold $\overline{q_i}$, the probability of product improvement during the campaign actually increases with q_i provided that the complexity of the product does not increase too much with the additional features (i.e., $b(q_i + q_u) < 0.5$). This is because the customer's utility increases with q_i, and hence the customer is more likely to pledge and make comments to induce the creator to improve the product. When q_i is above $\overline{q_i}$, the creator does not improve the product during the campaign. The intuition is as follows. The creator decides whether to improve the product by comparing how much improvement would increase the chance of campaign success with the additional production and investment cost. When q_i is high, improving the product does not sufficiently increase the customer's likelihood of pledging (and hence campaign success) to justify the large increase in the creator's cost. Therefore, the creator improves the product during the campaign only when q_i is below a certain threshold.[14]

Akin to Proposition 1, Proposition 2 analyzes the impact of q_i on $\mathbb{P}(improve)$ for each campaign. Thus, to align our theoretical prediction better with our empirical analysis, we analyze the average impact by conducting a numerical analysis as described in §2.2. Based on Proposition 2 and our numerical analysis illustrated in Figure 3, we establish the following hypothesis (hereafter H2).

Hypothesis 2. *As the initial number of product features increases, the likelihood that the product will be improved during the campaign first increases and then decreases.*

2.4. Probability of Campaign Success

Using Lemma 1, we next answer our third research question (Q3). The following proposition characterizes the impact of the initial number of features q_i on $\mathbb{P}(success)$.

Proposition 3. *Let $\overline{q_i}$ be the threshold defined in Proposition 1. When $q_i \le \overline{q_i}$, $\mathbb{P}(success)$ is increasing in $q_i \in (0, \frac{0.5}{b} - q_u)$ and decreasing in $q_i \in (\frac{0.5}{b} - q_u, \overline{q_i})$. When $q_i > \overline{q_i}$, $\mathbb{P}(success)$ is increasing in $q_i \in (\overline{q_i}, \frac{0.5}{b})$ and decreasing in $q_i \in (\frac{0.5}{b}, \infty)$.*

Proposition 3 shows that regardless of anticipating product improvement during the campaign or not (i.e., $q_i \le \overline{q_i}$ or $q_i > \overline{q_i}$, respectively), the probability of campaign success increases with the initial number of features q_i as long as $bq_f < 0.5$. Otherwise, the probability of campaign success decreases with q_i. This is because both customers' utilities increase with q_i and hence customers are more likely to pledge as long as the product is not too complex.

In a similar fashion to Proposition 1 and Proposition 2, Proposition 3 analyzes the impact of q_i on $\mathbb{P}(success)$ for each campaign, and so we analyze the average impact by conducting a numerical analysis as described in §2.2. Based on Proposition 3 and our numerical analysis illustrated in Figure 3, we establish the following hypothesis (hereafter H3).

Hypothesis 3. *As the initial number of product features increases, the likelihood of campaign success first increases and then decreases.*

Table 1 summarizes the positive and negative effects of an increase in q_i on different metrics.

[14] Even when the marginal cost of production is constant (i.e., cost of production is $c \cdot q_f$), this result continues to hold because as q_i increases, improving the product leads to a smaller increase in the likelihood of the customer pledging.

Table 1 How an increase in q_i affects $E[\#of comments]$, $\mathbb{P}(comment)$, $\mathbb{P}(improve)$, $\mathbb{P}(success)$.

	Positive Effects	Negative Effects
$\mathbb{P}(comment)$	Higher value of the product	Higher complexity
$E[\#of comments]$	Higher value of the product, more comments per customer (if n > 0)	Higher complexity, less comments per customer (if n < 0)
$\mathbb{P}(improve)$	Higher value of the product	Higher production and investment cost
$\mathbb{P}(success)$	Higher value of the product	Higher complexity

3. Empirical Models and Analysis

We test our hypotheses using a unique data set collected from Kickstarter, a crowdfunding platform that enables creators to launch reward-based, all-or-nothing campaigns. To launch a campaign on Kickstarter, each creator prepares a campaign page that includes a textual description of the product along with supporting materials such as pictures. On the campaign page, the creator also specifies a funding goal and the pledge price required to receive the product as a reward. Once the campaign has been launched, a customer who arrives on the campaign page reads the product description to decide whether or not to pledge. As we discuss in §1, if making a pledge, the customer can make a comment on the campaign page to induce the creator to improve the product further (customers cannot make comments before pledging; see Kickstarter 2021g). If the product is then improved, the creator revises the product description accordingly. After the campaign ends, customers cannot pledge and the creator cannot revise the product description. As we discuss in §2, if the total amount pledged at the end exceeds the funding goal, then the campaign is successful; otherwise, the campaign fails and customers are refunded. In the remainder of this section, we describe the sample, variables, empirical models, empirical results, and robustness analyses.

3.1. Sample

Consistent with our theoretical model, we focus on 21,768 campaigns for physical products in the technology and design categories launched on Kickstarter between July 2013 and February 2016.[15] The sample contains 6,488 successful campaigns, 12,111 failed campaigns, and 3,169 canceled campaigns.[16] We exclude 388 campaigns that are not suitable for textual analysis (e.g., non-English campaigns, see §3.2), so the final sample contains 21,380 campaigns.

[15] These campaigns constitute the majority of the technology and design campaigns, and include product subcategories such as camera equipment, hardware, and product design, but not software, web, and graphic design. We restrict our attention to these physical product categories because a campaign for a non-physical product may have a different production cost structure in the theoretical model.

[16] Campaigns can be canceled due to intellectual property disputes or at the discretion of creators (Kickstarter 2021b). Although we exclude canceled campaigns in our main empirical analyses, our empirical results continue to hold when we treat them as failed campaigns (see §3.4).

3.2. Dependent, Explanatory, and Control Variables

Dependent Variables: Customer Feedback, Product Improvement, and Campaign Success. We operationalize customer feedback in two ways. First, to test H1, we need a variable that measures whether the campaign received any comments up until the campaign reached the funding goal or the campaign ended, whichever was earlier. Therefore, we first generate a binary variable for customer feedback, existence of comments EC_k, for each campaign k. Specifically, $EC_k = 1$ if the campaign received any comments; otherwise, $EC_k = 0$. In our sample, at least one comment was received during 54% of all campaigns and 81% of successful campaigns. Furthermore, we measure the number of comments NC_k received by campaign k up until the campaign reached the funding goal or the campaign ended, whichever was earlier.

To test H2, we need a variable that measures whether a creator improved the product during a campaign or not. Hence, we generate a binary variable, product improvement I_k, for each campaign k, which represents whether the final number of features q_{fk} of the product is higher than its initial number of features q_{ik}. (We explain how we construct q_{ik} and q_{fk} when we discuss the number of features below.) Specifically, $I_k = 1$ if $q_{fk} > q_{ik}$; otherwise, $I_k = 0$. In our sample, 26% of all products and 43% of products in successful campaigns were improved during their campaigns.[17]

To test H3, we create a binary variable, campaign success S_k, for each campaign k which represents whether the total amount pledged P_k at the end of the campaign was greater than or equal to the funding goal G_k (e.g., Mollick 2014, Wei et al. 2020). Specifically, $S_k = 1$ if $P_k \geq G_k$; otherwise, $S_k = 0$. This measure is important for creators to evaluate their success, and it is also consistent with how Kickstarter evaluates campaigns to analyze the performance of the platform (Kickstarter 2021f). In our sample, 35% of campaigns were successful (excluding canceled campaigns).

Number of Product Features. Following the prior literature (e.g., Tirunillai and Tellis 2014, Toubia et al. 2019), we use latent Dirichlet allocation (LDA) (Blei et al. 2003) to create a proxy for the number of features by extracting topics and their weights in each description. Our approach is as follows. We train the LDA model on 42,564 initial and final descriptions of 21,380 campaigns in our sample. (We discuss the details of the LDA model in §EC.8 in the Online Appendix.) In line with the prior literature (e.g., Blei et al. 2003, Griffiths and Steyvers 2004), we consider a topic to be present in a description if it is associated with at least ten words. (Our empirical results continue to hold when we use different thresholds instead of ten words; see §3.5.) As campaigns vary in their description lengths (in words), for each campaign we first calculate "10/the description length" as a threshold, and then compare each topic weight with this threshold to determine whether the

[17] To make our empirical model consistent with our theoretical model, we use a binary measure of product improvement. In §EC.3 in the Online Appendix, we show that our result for H2 continues to hold when we use alternative definitions of product improvement, e.g., $I_k = q_{fk} - q_{ik}$.

topic is present in the description. Thus, in each campaign k, we measure the product's initial number of features q_{ik} as the number of topics in its initial description and the product's the final number of features q_{fk} as the number of topics in its final description.[18]

For example, the initial and final numbers of features in the HAIZE navigation device discussed in §1 are calculated as 14 and 19, consistent with the fact that the product was improved during its campaign. (See Figure EC.8.1 in the Online Appendix for the initial and final descriptions of this product.) In our sample, the average initial number of features is 9.02 and the average final number of features is 9.59. As we need both q_{ik} and q_{fk} to calculate I_k, we exclude 196 campaigns which have only either an initial or a final description (see §EC.8 in the Online Appendix).

To confirm that the number of features our LDA model predicts is a good proxy for the number of features that products have, we use data of 4,153 campaigns from another crowdfunding platform (Indiegogo), and provide evidence for the relationship between the number of features that our LDA model predicts for each product and the product's development level reported by the creator (see §EC.9 in the Online Appendix). As products that are more developed tend to have more features (e.g., Althuizen and Chen 2021), our additional analysis suggests that our LDA model generates a good proxy for the number of features that products have.

Control Variables. In our empirical models, we include several controls for campaign and creator characteristics. Specifically, we control for the *category* of each campaign (i.e., technology or design; we set design as the base category in empirical models) because the initial number of product features can differ across categories. Also, we control for each campaign's funding *goal* (natural logarithm of goal in US dollars) and *duration* (in days) (e.g., Mollick 2014, Blaseg et al. 2020). These variables enable us to control for the scale of a project because, for example, we expect the goal to be higher and/or the duration of the campaign to be longer for a larger scale project. Additionally, we control for the median *pledge price* (in US dollars) of each campaign and the *delivery time* (the number of months between the last delivery date and the end of the campaign), which are also linked to the scale of the project.[19]

In our analysis, we derive the number of product features from their textual descriptions. Because these descriptions can also include videos and pictures as well as a section in which creators discuss various risks associated with their campaigns, we control for the number of *videos*, the number of *pictures*, and *risk-section length* (in words) (e.g., Blaseg et al. 2020).

[18] To verify that the LDA model works well, we randomly select 20 products and confirm that their descriptions are consistent with features identified in these descriptions. Also, we manually compare the final and initial descriptions of these randomly selected products, and observe that revisions to each description during the campaign are consistent with additional features identified by our LDA model. In only one case, although the product was improved during the campaign, our LDA model could not identify the additional feature.

[19] Our additional analysis indicates that the median pledge price is a good proxy for the pledge price of the product in each campaign, but our empirical results hold when we control for the mean pledge price in each campaign.

We include additional control variables to take account of various creator-related factors that may affect our dependent variables. First, we control for *creator experience* in terms of the number of previous campaigns launched by a creator. Second, we control for whether or not the creator is an *individual*, which we define as follows. If the majority of personal pronouns in the product description are singular and the creator is not an organization with a legal name (e.g., Ltd, Inc), then the creator is an individual; if not, the creator is not an individual.

Finally, we control for the average level of *competition* during each campaign as follows. For each category and each day, we calculate the number of concurrent campaigns and the number of new customers.[20] We then divide these two variables to obtain the "campaign–customer" ratio. Then, for each campaign, we control for the average campaign–customer ratio during the campaign.

3.3. Model Specification

Our empirical strategy relies on probit models, an ordinary least-squares (OLS) regression model, and an instrumental variable (IV) approach to address potential endogeneity concerns. In §3.6, we also extend our models to capture the relationship between our dependent variables.

Probit Models. Because our dependent variables—existence of comments EC_k, product improvement I_k and campaign success S_k—are binary, we use probit models to test H1, H2, and H3. Let X_k be the vector of control variables for campaign k. First, to test the nonlinear (first increasing and then decreasing) relationship between the initial number of features q_{ik} on existence of comments EC_k, we include both q_{ik} and $(q_{ik})^2$, and obtain the following Probit Model 1:

$$P(EC_k) = \Phi(\alpha_0 + \alpha_1 q_{ik} + \alpha_2 (q_{ik})^2 + \alpha_X X_k + v_k), \tag{R1}$$

where Φ is the cumulative distribution function of the standard normal distribution. Similarly, to test the nonlinear relationship between q_{ik} on product improvement I_k and campaign success S_k, we use the same model by replacing EC_k with product improvement I_k and campaign success S_k, and obtain the following Probit Models 2 and 3, respectively:

$$P(I_k) = \Phi(\beta_0 + \beta_1 q_{ik} + \beta_2 (q_{ik})^2 + \beta_X X_k + u_k), \tag{R2}$$

$$P(S_k) = \Phi(\gamma_0 + \gamma_1 q_{ik} + \gamma_2 (q_{ik})^2 + \gamma_X X_k + w_k). \tag{R3}$$

Linear Regression. To test the impact of the initial number of features q_{ik} on the number of comment(s) NC_k in, we use the following regression model:

$$NC_k = \theta_0 + \theta_1 q_{ik} + \theta_2 (q_{ik})^2 + \theta_X X_k + r_k. \tag{R4}$$

[20] On December 19, 2013, the number of new customers is zero in both categories due to a server error, so we replace the number of new customers on this day with the average number of customers in each category.

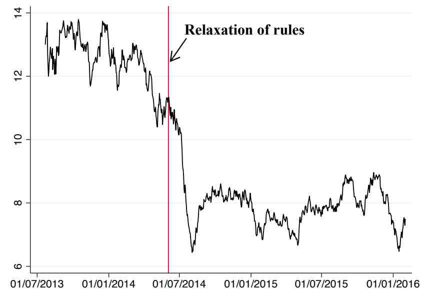


Figure 4 30-day moving average of initial number of product features by campaign start dates.

To test the (possible) nonlinear relationship between q_{ik} and NC_k, we include both q_{ik} and $(q_{ik})^2$.

Instrumental Variable. Although we control for campaign and creator characteristics, there may still be unobserved factors that can simultaneously affect the initial number of product features q_{ik} in campaign k and our outcome variables. To address this problem and any potential measurement errors, we use an exogenous shock to the number of product features in new campaigns. Specifically, on June 3, 2014, Kickstarter introduced a new policy which simplified its rules and allows creators to launch campaigns whenever they feel ready without forcing them to satisfy eligibility requirements for launching a campaign and without prior review by Kickstarter (Kickstarter 2021d). As Figure 4 illustrates, this exogenous shock leads to a substantial decrease in the average initial number of product features. Using this exogenous shock as an instrument allows us to isolate the impact of the initial number of features on our outcome variables.[21]

To use this instrument in our IV models, we create a binary variable, *before relaxation of rules* B_k, where $B_k = 1$ if campaign k is launched before June 3, 2014, and $B_k = 0$ otherwise. As we focus on campaigns launched between July 2013 and February 2016, we have comparable time periods before and after the instrument. This also enables us to avoid a confounding event on February 2, 2016, when Kickstarter started to select curated campaigns (Kickstarter 2021c). Note that the $F-$statistic in the regression of the initial number of product features on B_k is 900.10, which indicates that our IV satisfies the relevance condition.

IV Models. In our empirical analysis, as we first aim to analyze the nonlinear relationship between an endogenous explanatory variable and binary dependent variables, it would be problematic to use a standard two-stage least squares (2SLS) approach. Thus, we implement the IV using a control function (CF) approach (Wooldridge 2010, 2015), a two-step approach that allows us to condition

[21] A concern may be that the relaxation of rules simultaneously leads to an increase in the number of campaigns and thereby reduces each campaign's likelihood of success. This would violate the exclusion restriction by which an instrument cannot affect an outcome variable directly but only through the instrumented explanatory variable. To satisfy the exclusion restriction, we include the level of competition as an additional control variable. The instrument is then independent of campaign success, conditional on the explanatory variable and the level of competition.

out the variation in unobserved variables that depends on the endogenous variable, and hence the remaining variation in the endogenous variable is independent of the error (Petrin and Train 2010).

To test H1 with the IV model, we use the following procedure as explained in Wooldridge (2015) (see Chan et al. 2020 for an example in the operations management literature). In the first stage, we regress the initial number of product features q_{ik} on the instrumental variable B_k and control variables in an OLS regression model. We then use the predicted residuals $\hat{u}_k$ of the first-stage regression in the second-stage probit model whose dependent variable is existence of comments EC_k (Wooldridge 2015). Because $\hat{u}_k$ is an estimate from the first stage, which adds extra variation in the second stage (Petrin and Train 2010), we also use a nonparametric bootstrap to obtain valid standard errors in the second stage (Wooldridge 2010, 2015). For H1, we obtain the IV model (IV Model 1a) with the following first- and second-stage regressions:

$$q_{ik} = \alpha_0 + \alpha_1 B_k + \alpha_X X_k + u_k, \text{ and} \tag{R5}$$

$$P(EC_k) = \Phi(\beta_0 + \beta_1 q_{ik} + \beta_2 (q_{ik})^2 + \beta_3 \hat{u}_k + \beta_4 (\hat{u}_k)^2 + \beta_X X_k + v_k). \tag{R6}$$

To test the nonlinear relationship between q_{ik} and EC_k, we include both $(q_{ik})^2$ and $(\hat{u}_k)^2$ in the second-stage regression (Wooldridge 2015, page 437).[22] Similarly, to test H2 and H3, we obtain IV Model 2 and IV Model 3 by keeping the first stage (R5) and replacing EC_k in the second stage (R6) with I_k and S_k, respectively, as follows:

$$P(I_k) = \Phi(\gamma_0 + \gamma_1 q_{ik} + \gamma_2 (q_{ik})^2 + \gamma_3 \hat{u}_k + \gamma_4 (\hat{u}_k)^2 + \gamma_X X_k + z_k), \text{ and} \tag{R7}$$

$$P(S_k) = \Phi(\theta_0 + \theta_1 q_{ik} + \theta_2 (q_{ik})^2 + \theta_3 \hat{u}_k + \theta_4 (\hat{u}_k)^2 + \theta_X X_k + t_k). \tag{R8}$$

To test the (possible) nonlinear relationship between q_{ik} and NC_k with the IV model, we follow a similar procedure, by replacing the probit model in the second-stage with an OLS regression model whose dependent variable is the number of comments NC_k. Thus, we obtain the following IV Model 1b by keeping the first stage (R5) and replacing EC_k in the second stage (R6) of IV Model 1a with NC_k as follows:

$$NC_k = \omega_0 + \omega_1 q_{ik} + \omega_2 (q_{ik})^2 + \omega_3 \hat{u}_k + \omega_4 (\hat{u}_k)^2 + \omega_X X_k + w_k. \tag{R9}$$

3.4. Main Results

Tables 2 and 3 show the descriptive statistics and correlations (we exclude 3,011 canceled campaigns). We report no issue of multicollinearity. All main results are presented in Table 4. The

[22] This approach helps us avoid the forbidden regression problem in our model because we do not directly plug predicted values of q_{ik} from the first stage in the nonlinear second-stage regression (cf. Angrist and Pischke 2009). Instead, we implement a control function approach, which was developed as a solution to the forbidden regression problem (Wooldridge 2010, 2015, Petrin and Train 2010), and hence we use predicted residuals $\hat{u}_k$.

effects of control variables on our outcome variables in both models are as expected. For example, compared to teams, individual creators are less likely to improve their products and their campaigns are less likely to be successful.

Table 2 Descriptive statistics of variables in the empirical models ($n = 18,173$).

Variables	Mean	Standard deviation	Minimum	Maximum
Existence of comment(s)	0.54	0.50	0	1
Number of comments	6.54	24.48	0	1,625
Product improvement	0.26	0.44	0	1
Campaign success	0.35	0.48	0	1
Initial number of features	9.02	8.11	0	48
Goal (ln)	9.48	1.64	-0.28	18.52
Duration	34.63	10.5	1	61
Pledge price	174.95	535.09	0	10000
Delivery time	4.27	5.03	0	70
Videos	0.26	0.83	0	26
Pictures	11.08	12.11	0	119
Risk-section length	141.64	119.95	8	4981
Creator experience	0.18	0.78	0	21
Individual	0.31	0.46	0	1
Competition	0.16	0.04	0.05	0.54
Before relaxation of rules	0.22	0.41	0	1

Note. The minimum value of the natural logarithm of the goal (in US dollars) is negative because the minimum value of the goal is one Canadian dollar. Our empirical results continue to hold when we exclude 324 campaigns where the goal (in US dollars) is smaller than the 1st percentile ($168) or larger than the 99th percentile ($500,000).

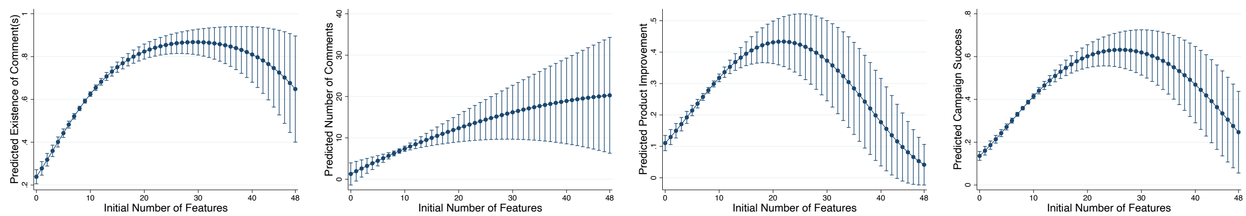
Table 3 Correlation matrix for variables in the empirical models ($n = 18,173$).

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) *Existence of comment(s)*	1.00														
(2) *Number of comments*	0.25***	1.00													
(3) *Product improvement*	0.25***	0.15***	1.00												
(4) *Campaign success*	0.40***	0.22***	0.28***	1.00											
(5) *Initial number of features*	0.30***	0.16***	0.18***	0.21***	1.00										
(6) *Goal* (ln)	0.13***	0.17***	0.07***	-0.22***	0.25***	1.00									
(7) *Duration*	0.06***	0.05***	0.04***	-0.05***	0.05***	0.22***	1.00								
(8) *Pledge price*	-0.05***	0.00	0.01	-0.04***	0.04***	0.18***	0.02***	1.00							
(9) *Delivery time*	-0.03***	0.03***	0.01*	-0.08***	0.05***	0.26***	0.10***	0.10***	1.00						
(10) *Videos*	0.10***	0.06***	0.05***	0.04***	0.18***	0.11***	0.04***	0.01	0.01*	1.00					
(11) *Pictures*	0.37***	0.22***	0.19***	0.32***	0.53***	0.17***	0.08***	0.00	0.00	0.18***	1.00				
(12) *Risk-section length*	0.15***	0.08***	0.09***	0.06***	0.36***	0.20***	0.03***	0.03***	0.07***	0.09***	0.21***	1.00			
(13) *Creator experience*	0.04***	0.02**	0.01*	0.17***	-0.01	-0.18***	-0.03***	-0.03***	-0.06***	0.00	0.04***	-0.03***	1.00		
(14) *Individual*	-0.16***	-0.09***	-0.09***	-0.12***	-0.14***	-0.18***	-0.06***	-0.02***	0.01	-0.06***	-0.21***	-0.08***	0.00	1.00	
(15) *Competition*	-0.09***	-0.03***	-0.06***	-0.09***	-0.14***	0.01	0.06***	0.04***	0.05***	0.02***	-0.09***	-0.04***	0.02**	0.01*	1.00

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We first discuss the results of the IV models (see Table 4). The second stage of IV Model 1a (column (6) in Table 4) shows the results for H1 regarding the existence of comment(s). While the coefficient of the initial number of features q_{ik} is positive and significant ($\beta_1 = 0.14$, $p < 0.01$), the coefficient of $(q_{ik})^2$ is negative and significant ($\beta_2 = -0.002$, $p < 0.01$). We also calculate the turning point as $q_{ik} = 30.69$ and its 95% confidence interval as $(26.08, 38.27)$. Both the turning point and its confidence interval are within the data range (e.g., Haans et al. 2016, Tan and Netessine 2019). As Figure 5(a) illustrates, this result indicates that as the initial number of features increases, the likelihood of the existence of comments first increases and then decreases, supporting H1.

The second stage of IV Model 1b (column (7) in Table 4) shows the results for the number of comments. While the coefficient of the initial number of features q_{ik} is positive and significant ($\omega_1 = 0.628$, $p < 0.01$), the coefficient of $(q_{ik})^2$ is insignificant. As Figure 5(b) illustrates, this result indicates that as the initial number of features increases, the number of comments always increases. This result is consistent with a positive-valued n in our theoretical model (see Proposition 1(b)),



(a) IV Model 1a (b) IV Model 1b (c) IV Model 2 (d) IV Model 3

Figure 5 Predicted likelihood of existence of comment(s), predicted number of comments, predicted likelihood of product improvement, and predicted likelihood of campaign success.

indicating that the number of comments that a pledging customer makes increases with the initial number of features because more features can stimulate more ideas.

The second stage of IV Model 2 (column (8) in Table 4) shows the results for H2 regarding product improvement. While the coefficient of the initial number of features q_{ik} is positive and significant ($\gamma_1 = 0.103$, $p < 0.01$), the coefficient of $(q_{ik})^2$ is negative and significant ($\gamma_2 = -0.002$, $p < 0.01$). We also calculate the turning point as $q_{ik} = 23.12$ and its 95% confidence interval as $(20.35, 25.76)$. Both the turning point and its confidence interval are within the data range. As Figure 5(c) illustrates, this result indicates that as the initial number of features increases, the likelihood of product improvement first increases and then decreases, supporting H2.

The second stage of IV Model 3 (column (9) in Table 4) shows the results for H3 regarding campaign success. While the coefficient of the initial number of features q_{ik} is positive and significant ($\theta_1 = 0.129$, $p < 0.01$), the coefficient of $(q_{ik})^2$ is negative and significant ($\theta_2 = -0.002$, $p < 0.01$). We also calculate the turning point as $q_{ik} = 26.01$ and its 95% confidence interval as $(21.35, 30.34)$. Both the turning point and its confidence interval are within the data range. As Figure 5(d) illustrates, this result shows that as the initial number of features increases, the likelihood of campaign success first increases but then decreases, supporting H3.

Columns (1), (3), and (4) in Table 4 also show that the results of probit models are consistent with the results of IV models. Also, as in IV Model 1b, in the regression model, the coefficient of the initial number of features q_{ik} is positive but not significant (see column (2) in Table 4). Finally, in the second stage of all IV models, the coefficients of residuals obtained from the first-stage model are negative and significant. Significant residuals confirm a possible endogeneity problem and support our use of an IV (Wooldridge 2010).

3.5. Robustness Analyses

To check the robustness of our empirical results, we run spline regressions, which use knots to capture the different impact of an explanatory variable for different intervals (e.g., Kesavan et al.

2014). We try various spline regressions for the second-stage estimations in both IV Models 2 and 3. We find that the coefficient of the first spline is positive and significant and that the coefficient of the second spline is negative and significant (see Table EC.10.1 in §EC.10 in the Online Appendix), supporting our results about H2 and H3. Although the coefficient of the second spline of IV Model 1a is negative, it is not significant. In line with our main models, this indicates that the number of product features has a slightly weaker curvilinear relationship with commenting than with the other dependent variables.

We use alternative definitions of product improvement to establish robustness. First, for each campaign k, we measure product improvement as $I_k = 1$ if $q_{fk} \neq q_{ik}$; otherwise, $I_k = 0$. In doing so, in addition to 4,717 campaigns where $q_{fk} > q_{ik}$, we classify 1,168 campaigns where $q_{fk} < q_{ik}$ as campaigns where the product is improved during the campaign. Also, we use an alternative definition of product improvement where we identify additional 394 campaigns where $q_{fk} = q_{ik}$ but where there is a change in the existing product features during the campaign. We classify them as $I_k = 1$ in addition to campaigns where $q_{fk} > q_{ik}$ or $q_{fk} < q_{ik}$. As shown in Table EC.5.1 in §EC.5 in the Online Appendix, our empirical results continue to hold in both cases.

We show the robustness of our empirical findings in the following cases (see Online Appendix EC.10 for details). First, to have equal time periods before and after the instrument, we exclude campaigns launched after April 28, 2015, and establish robustness by analyzing 11,764 campaigns. As presented in Table EC.10.2, the only difference is that the positive impact of the initial number of features on the number of comments becomes insignificant in this case. Second, we treat canceled campaigns as failed campaigns, and establish robustness by analyzing 21,184 campaigns. Third, as we explain in §EC.8 in the Online Appendix, we train the LDA model with 50 topics, in line with the literature. Our results continue to hold when we set the number of topics to 20% above or below 50 in the LDA model and when we set the threshold to 20% above or below 10 words when counting the number of topics in each description. Fourth, when testing H1 and H2, we control for the average competition in the first week of each campaign rather than the average competition during the campaign in order to avoid a timing problem, and we show that our results hold.

Table 4 Results of main models.

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Regression equation	(R1)	(R4)	(R2)	(R3)	(R5)	(R6)	(R9)	(R7)	(R8)
Model	Probit Model 1	Regression Model	Probit Model 2	Probit Model 3	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Existence of comment(s)	Number of comments	Product improvement	Campaign success	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features	.07***	.143	.072***	.088***		.14***	.628***	.103***	.129***
	(.004)	(.087)	(.004)	(.004)		(.009)	(.136)	(.007)	(.009)
Initial number of features^2	-.002***	-.002	-.002***	-.002***		-.002***	-.003	-.002***	-.002***
	(0)	(.003)	(0)	(0)		(0)	(.004)	(0)	(0)
Category: Technology	-.011	2.534***	.059***	.025	1.004***	-.083***	2.011***	.026	-.016
	(.023)	(.443)	(.021)	(.024)	(.092)	(.023)	(.433)	(.023)	(.025)
Goal (ln)	.042***	1.911***	-.006	-.331***	.579***	.003	1.637***	-.023***	-.354***
	(.006)	(.151)	(.007)	(.008)	(.031)	(.008)	(.177)	(.008)	(.009)
Duration	.003***	.012	.003***	-.002**	-.019***	.005***	.02	.003***	-.001
	(.001)	(.02)	(.001)	(.001)	(.004)	(.001)	(.021)	(.001)	(.001)
Pledge price	0***	-.001***	0	0***	0	0***	-.001***	0	0***
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Delivery time	-.01***	-.015	.002	-.005**	.011	-.011***	-.017	.002	-.005**
	(.002)	(.041)	(.002)	(.002)	(.011)	(.002)	(.041)	(.002)	(.002)
Videos	.031***	-.022	.015	-.005	.696***	-.014	-.349	-.006	-.032**
	(.012)	(.28)	(.013)	(.013)	(.091)	(.014)	(.287)	(.015)	(.013)
Pictures	.035***	.359***	.013***	.033***	.296***	.015***	.216***	.004*	.022***
	(.002)	(.04)	(.001)	(.001)	(.007)	(.003)	(.046)	(.002)	(.003)
Risk-section length	0***	.001	0***	0***	.015***	-.001***	-.006***	0	0**
	(0)	(.001)	(0)	(0)	(.001)	(0)	(.002)	(0)	(0)
Creator experience	.066***	1.046***	.02	.257***	.203***	.06***	1.012***	.018	.257***
	(.015)	(.18)	(.014)	(.025)	(.06)	(.015)	(.189)	(.014)	(.025)
Individual	-.212***	-1.043**	-.14***	-.357***	-.04	-.213***	-1.04**	-.14***	-.36***
	(.021)	(.474)	(.023)	(.026)	(.102)	(.021)	(.465)	(.023)	(.026)
Competition	-1.568***	-14.09***	-1.274***	-2.056***	-2.543**	-.119	-3.72	-.595*	-1.163***
	(.251)	(4.219)	(.285)	(.265)	(1.253)	(.327)	(5.391)	(.318)	(.329)
Before relaxation of rules					3.513***				
					(.128)				
Residuals						-.07***	-.498***	-.031***	-.04***
						(.009)	(.13)	(.006)	(.009)
Residuals×Residuals						.001**	.003	0	0**
						(0)	(.007)	(0)	(0)
Constant	-.796***	-15.571***	-1.04***	2.261***	-2.292***	-.912***	-16.297***	-1.095***	2.192***
	(.075)	(1.796)	(.088)	(.081)	(.364)	(.078)	(1.917)	(.089)	(.084)
Wald χ^2	2166.15	1128.07	1391.76	3415.19	9001.21	2502.92	2031.22	1480.71	3602.76
R^2 or Pseudo R^2	.151	.072	.056	.204	.399	.154	.073	.057	.206
Observations	18,173	18,173	18,173	18,173	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table 5 Impact of customer feedback and product improvement on campaign success

Column	(1)	(2)	(3)	(4)	(5)	(6)
Regression equation	(R10)	(R11)	(R12)	(R13)	(R12) with IV	(R13)
Model	First Stage of IV Models	First Stage of IV Model with Biprobit	Second Stage of IV Model with Biprobit	Second Stage of IV Model with Biprobit	Second Stage of IV Model with Biprobit	Second Stage of IV Model with Biprobit
	Initial number of features	Number of comments (log)	Product improvement	Campaign success	Product improvement	Campaign success
Before relaxation of rules	3.019***					
	(.168)					
Customers' previous pledges (log)	.11*	.009**	.009	-.343***	.004	-.383***
	(.063)	(.005)	(.018)	(.02)	(.018)	(.026)
Initial number of features			.068***	.11***	.068***	.114***
			(.013)	(.012)	(.014)	(.013)
Initial number of features^2			-.002***	-.002***	-.002***	-.002***
			(0)	(0)	(0)	(0)
Residuals from (1)			-.01	-.026**	-.01	-.029**
			(.012)	(.011)	(.012)	(.012)
Residuals from (1)^2			0	.001**	0	.001**
			(0)	(0)	(0)	(0)
Creator's improvement experience					1.098***	
					(.135)	
Customers' previous comments (log)		1.273***				
		(.1)				
Number of comments (log)			1.004***	1.589***	1.005***	1.643***
			(.129)	(.152)	(.13)	(.165)
Number of comments (log)^2			-.268***	-.548***	-.275***	-.586***
			(.045)	(.052)	(.047)	(.057)
Residuals from (2)			0	.753***	.009	.858***
			(.12)	(.149)	(.12)	(.161)
Residuals from (2)^2			.156***	.381***	.147***	.403***
			(.052)	(.051)	(.053)	(.055)
Product improvement				-1.017***		-.797***
				(.041)		(.134)
ρ				.880***		.771***
				(.021)		(.068)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.119***	-.134***	-1.056***	3.085***	-1.025***	3.435***
	(.463)	(.027)	(.1)	(.129)	(.103)	(.202)
Wald χ^2	7265.24	4024.14	10052.76	10052.76	9371.99	9371.99
R^2 or Pseudo R^2	.414	.296				
Observations	13,568	13,568	13,568	13,568	13,568	13,568

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

3.6. Additional Analysis with Biprobit Models

Until now, following up on our main research questions and hypotheses, we focused on the initial number of product features and its effects on customer feedback, product improvement, and campaign success. However, there can be dependencies between customer feedback, product improvement, and campaign success.[23] In this section, we explicitly consider possible interplay among our dependent variables. Specifically, we analyze the impact that the initial number of features and the number of comments have on product improvement and campaign success while also capturing the impact of product improvement on campaign success.

For this analysis, we need a new IV for the number of comments. For each campaign k, we calculate *customers' previous comments* (PC_k) which is the average number of comments that the focal campaign's customers made in previous campaign(s) conditional on having pledged. (Note that we control for the average number of previous campaigns that those customers pledged in, log_{10}*(customers' previous pledges+1)*.) *Customers' previous comments* represents the customers' tendency to make comments and should significantly affect the number of comments that those customers make in campaign k while having no direct impact on the creator's product improvement decision in the same campaign k or on the success of campaign k. In constructing our instrumental variable, to deal with outliers, we calculate the instrumental variable as $log_{10}(PC_k+1)$ (e.g., Koning et al. 2022). (Our results are similar when we use the inverse hyperbolic sine transformation (e.g., Bulte et al. 2017).) Note that this data set is available for 13,435 campaigns ending before August 2015 because pledge information is not available on Kickstarter after August 2015.

As we analyze the impact of both q_{ik} and the number of comments NC_k, in addition to the first-stage regression (R5) of our main model in §3.4, we have another first-stage regression where we regress the number of comments $log_{10}(NC_k+1)$ on the instrumental variable PC_k and control

[23] We thank the review team for pointing this out.

variables in an ordinary least-squares regression model (R11). (To deal with non-convergence issues stemming from outliers, we use $log_{10}(NC_k + 1)$ instead of NC_k. Our results are similar when we use the inverse hyperbolic sine transformation.) We then use the predicted residuals $\hat{u}_k$ and $\hat{v}_k$ of the first-stage regressions in two second-stage models whose dependent variables are product improvement I_k (R12) and campaign success S_k (R13). To test the impact of product improvement I_k on campaign success S_k, we also include I_k on the right-hand side of equation (R13), and we estimate models (R12) and (R13) simultaneously in a recursive bivariate probit model (biprobit) (e.g., Greene 2018, Liu et al. 2019, Freeman et al. 2021). So, we obtain the following IV model with two first-stage and two second-stage regressions (i.e., CF Biprobit, Lin and Wooldridge 2019):

$$q_{ik} = \alpha_0 + \alpha_1 B_k + \alpha_X X_k + u_k, \tag{R10}$$

$$log_{10}(NC_k + 1) = \beta_0 + \beta_1 log_{10}(PC_k + 1) + \beta_X X_k + v_k, \tag{R11}$$

$$\begin{aligned} P(I_k) =& \Phi(\gamma_0 + \gamma_1 q_{ik} + \gamma_2 (q_{ik})^2 + \gamma_3 log_{10}(NC_k + 1) + \gamma_4 (log_{10}(NC_k + 1))^2 + \gamma_5 \hat{u}_k + \gamma_6 (\hat{u}_k)^2 + \gamma_7 \hat{v}_k \\ &+ \gamma_8 (\hat{v}_k)^2 + \gamma_X X_k + z_k), \end{aligned} \tag{R12}$$

$$\begin{aligned} P(S_k) =& \Phi(\theta_0 + \theta_1 q_{ik} + \theta_2 (q_{ik})^2 + \theta_3 log_{10}(NC_k + 1) + \theta_4 (log_{10}(NC_k + 1))^2 + \theta_5 I_k + \theta_6 \hat{u}_k + \theta_7 (\hat{u}_k)^2 \\ & \theta_8 \hat{v}_k + \theta_9 (\hat{v}_k)^2 + \theta_X X_k + t_k). \end{aligned} \tag{R13}$$

Table 5 shows the results. Columns (1) and (2) show the first-stage regressions (R10) and (R11). The second-stage regressions (R12) and (R13) are shown in columns (3) and (4) without an instrument for product improvement and in columns (5) and (6) with an instrument for product improvement.[24]

We derive the following results. First, as can be seen from columns (3) and (4) of Table 5, the impact of the initial number of features on the likelihood of product improvement ($\gamma_1 = 0.068$ and $\gamma_2 = -0.002$, $p < 0.01$) and on the likelihood of campaign success ($\theta_1 = 0.11$ and $\theta_2 = -0.002$, $p < 0.01$) continue to hold. Second, column (3) of Table 5 shows that the likelihood of product improvement increases with the number of comments ($\gamma_3 = 1.004$, $p < 0.01$), but too many comments can lead to a decrease in the likelihood of product improvement ($\gamma_4 = -0.268$, $p < 0.01$).[25] Third, column (4) of Table 5 shows that the likelihood of campaign success increases with the number of comments ($\theta_3 = 1.589$, $p < 0.01$), but too many comments can lead to a decrease in the likelihood of campaign success ($\theta_4 = -0.548$, $p < 0.01$). Columns (5) and (6) of Table 5 show that these results continue to hold when we use an IV for product improvement I_k.

[24] Biprobit models are identified without additional instruments due to their nonlinear transformation. To further strengthen identification (e.g., Greene 2018, Freeman et al. 2021), we use the additional instrument *creator's improvement experience* measured by the number of previous campaigns during which the creator improved the product.

[25] In §EC.6 in the Online Appendix, we also analyze the effect of a binary variable for comments (i.e., the existence of comments) as a robustness check and show that it has a positive impact on the likelihood of product improvement. Also, in §EC.7 in the Online Appendix, we analyze the mediating role of the number of comments and show that the likelihood of product improvement increases with the number of comments driven by the initial number of features.

Columns (4) and (6) of Table 5 interestingly show that the exogenous impact of product improvement on the likelihood of campaign success is negative and significant ($\theta_5 = -1.017$, $p < 0.01$ in column (4); and $\theta_5 = -0.797$, $p < 0.01$ in column (6)), while the estimated correlation coefficient that measures the correlation between the error terms z_k and t_k is positive and significant ($\rho = 0.880$, $p < 0.01$ in column (4); and $\rho = 0.771$, $p < 0.01$ in column (6)). These results suggest that although there are unobservable variables that increase the likelihood of product improvement and the likelihood of campaign success simultaneously, the likelihood of campaign success is not higher just because the product is improved during the campaign. We further investigate this result in §EC.7 in the Online Appendix by analyzing the impact of the initial number of features and the number of comments on campaign success through product improvements. We show that when the creator takes account of the initial number of features and customer feedback, product improvements increase the chance of campaign success.

4. Discussion and Conclusion

Crowdfunding is more than just an effective financing instrument for entrepreneurs. One of its key advantages, which has received cursory attention in the crowdfunding literature, is that it enables creators to improve their products in response to customer feedback (e.g., Mollick 2014). To take advantage of this opportunity, a creator may launch a crowdfunding campaign using a basic version of a product which contains fewer features, leaving room for further improvements with new features during the campaign. However, if the product appears to be underdeveloped, customers may not pledge and the campaign may therefore fail. With the aim of investigating this key trade-off, we studied how the initial number of product features affects customer feedback, product improvement, and campaign success.

Inspired by both crowdfunding practice and literature (e.g., Hu et al. 2015, Belavina et al. 2020, Chakraborty and Swinney 2021), we constructed a parsimonious game-theoretical model of reward-based, all-or-nothing crowdfunding which takes account of account customers' commenting decisions and the creator's product improvement decision. We then tested the theoretical predictions using original data from Kickstarter and obtained the following results. Although the likelihood that customers will comment first increases and then decreases with the initial number of product features, the number of comments increases with the initial number of features.

Furthermore, one might expect that products with more features are less likely to be improved as this will increase the creator's investment and production cost. However, we showed that this is only true when the product has many features. In contrast, when the product has few features, the likelihood of product improvement increases with the initial number of features, because customers are then more likely to pledge and to leave comments. We also showed that, as the initial number of features increases, the likelihood of campaign success first increases and then decreases.

Finally, we showed the robustness of our results by considering the interplay between customer feedback, product improvement, and campaign success. This analysis also helped us generate additional interesting results. First, customer feedback leads to a higher likelihood of product improvement and campaign success. Second, although the likelihood of product improvement and the likelihood of campaign success are positively correlated, this does not mean that a campaign is more likely to be successful just because the product is improved during the campaign. Rather, the creator should take account of the initial number of features and customer feedback when improving the product, because otherwise product improvements can harm campaign success.

Our results suggest that creators should avoid overdeveloping their products before campaigns because this can decrease the chance of campaign success due to the complexity of the product (e.g., Thompson et al. 2005). Furthermore, adding too many features before campaigns can also hinder the opportunity for creators to involve customers in product development during the campaign, which in turn also leads to a lower chance of campaign success.

As well as contributing to the crowdfunding literature, our results add to the NPD literature (cf. Krishnan and Ulrich 2001, Loch and Kavadias 2008), as we revisit the debate around flexible approaches in product development and the role of customer feedback. We point out that, unlike traditional product development approaches (e.g., Bhattacharya et al. 1998, Thomke and Reinertsen 1998), crowdfunding enables a creator to improve a product based on customer feedback before committing to production. Although this approach comes with a risk of campaign failure, our results suggest that it can still be better for the creator to leave scope to refine the product based on customer feedback during the campaign. Furthermore, our findings contribute to open innovation literature (e.g., Chesbrough 2003, Gambardella et al. 2017, Bogers et al. 2019) by demonstrating the benefit of moving from a closed innovation strategy where an entrepreneur fully completes development before launching a crowdfunding campaign towards an open innovation strategy where the entrepreneur incorporates external ideas from customers during a campaign. While the open innovation literature has considered limits to external search (West and Bogers 2014), we show that the level of initial in-house development significantly affects the availability of external search, as both low and high initial development levels reduce customer contributions.

As our paper is a first step towards analyzing creators' product development and improvement decisions in crowdfunding campaigns, it naturally has some limitations that provide opportunities for future research. First, when analyzing how the initial number of product features affects a campaign's chance of success, we measure this by comparing the funds raised at the end of the campaign to the funding goal. Although this measure is consistent with our theoretical model and appropriate for the purpose of our study, it would also be interesting to analyze the impact of the initial number of features on the likelihood of it being ultimately delivered to customers. This analysis would require a more comprehensive data set which includes product launch information. Also, as our focus is on the creator's product development and improvement decisions, we analyze the impact of the initial number of features on the number of customer comments. However, an interesting research direction would be to conduct the textual analysis of customer feedback. Finally, in our empirical analysis, we focus on the campaigns in the technology and design categories, which are two major categories on Kickstarter. Although this approach enable us to use the LDA model and obtain meaningful product features, a future study can focus on analyzing the impact of the initial number of features for other categories such as Games.

References

Agrawal, A., C. Catalini, A. Goldfarb. 2015. Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy* **24**(2) 253–274.

Allon, G., V. Babich. 2020. Crowdsourcing and crowdfunding in the manufacturing and services sectors. *Manufacturing & Service Operations Management* **22**(1) 102–112.

Althuizen, N., B. Chen. 2021. Crowdsourcing ideas using product prototypes: The joint effect of prototype enhancement and the product design goal on idea novelty. *Management Science* Forthcoming.

Angrist, J. D., J. S. Pischke. 2009. *Mostly harmless econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Babich, V., S. Marinesi, G. Tsoukalas. 2021. Does crowdfunding benefit entrepreneurs and venture capital investors? *Manufacturing & Service Operations Management* **66**(11) 4980–4997.

Belavina, E., S. Marinesi, G. Tsoukalas. 2020. Rethinking crowdfunding platform design: Mechanisms to deter misconduct and improve efficiency. *Management Science* **66**(11) 4980–4997.

Belleflamme, P., T. Lambert, A. Schwienbacher. 2014. Crowdfunding: Tapping the right crowd. *Journal of Business Venturing* **29**(5) 585–609.

Bellstam, G., S. Bhagat, J. A. Cookson. 2020. A text-based analysis of corporate innovation. *Management Science,* Forthcoming.

Bhaskaran, S., S. S. Erzurumlu, K. Ramachandran. 2020. Sequential product development and introduction by cash-constrained start-ups. *Manufacturing & Service Operations Management,* Forthcoming.

Bhattacharya, S., V. Krishnan, V. Mahajan. 1998. Managing new product definition in highly dynamic environments. *Management Science* **44**(11) 50–64.

Bi, G., B. Geng, L. Liu. 2019. On the fixed and flexible funding mechanisms in reward-based crowdfunding. *European Journal of Operational Research* **279**(1) 168–183.

Blaseg, D., C. Schulze, B. Skiera. 2020. Consumer protection on Kickstarter. *Marketing Science* **39**(1) 211–233.

Blei, D. M., A. Y. Ng, M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* **3** 993–1022.

Bogers, M., H. Chesbrough, S. Heaton, D. J. Teece. 2019. Strategic management of open innovation: A dynamic capabilities perspective. *California Management Review* **62**(1) 77–94.

Brown, C. L., G. S. Carpenter. 2000. Why is the trivial important? a reasons-based account for the effects of trivial attributes on choice. *Journal of Consumer Research* **26**(4) 372–385.

Bulte, E., R. Lensink, N. Vu. 2017. Do gender and business trainings affect business outcomes? experimental evidence from vietnam. *Management Science* **63**(9) 2885–2902.

Burtch, G., A. Ghose, S. Wattal. 2013. An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research* **24**(3) 499–519.

Burtch, G., D. Gupta, P. Martin. 2020. Referral timing and fundraising success in crowdfunding. *Manufacturing & Service Operations Management,* Forthcoming.

Castaño, R., M. Sujan, M. Kacker, H. Sujan. 2008. Managing consumer uncertainty in the adoption of new products: Temporal distance and mental simulation. *Journal of Consumer Research* **45**(3) 320–336.

Chakraborty, S., R. Swinney. 2019. Designing rewards-based crowdfunding campaigns for strategic (but distracted) contributors. *Working Paper, Duke University, Durham.*

Chakraborty, S., R. Swinney. 2021. Signaling to the crowd: Private quality information and rewards-based crowdfunding. *Manufacturing & Service Operations Management* **23**(1) 155–169.

Chan, T. H., J. Mihm, M. Sosa. 2020. Revisiting the role of collaboration in creating breakthrough inventions. *Manufacturing & Service Operations Management* Forthcoming.

Chemla, G., K. Tinn. 2020. Learning through crowdfunding. *Managemet Science* **66**(5) 1783–1801.

Chen, Y.-J., T. Dai, C. G. Korpeoglu, E. Körpeoğlu, O. Sahin, C. S. Tang, S. Xiao. 2020. OM Forum-innovative online platforms: Research opportunities. *Manufacturing & Service Operations Management* **22**(3) 430–445.

Chesbrough, H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press, Boston.

Choi, Y. R., M. Leÿvesque, D. A. Shepherd. 2008. When should entrepreneurs expedite or delay opportunity exploitation? *Journal of Business Venturing* **23** 333–355.

Cohen, M. A., J. Eliashberg, T.-H. Ho. 1996. New product development: The performance and time-to-market tradeoff. *Management Science* **42**(2) 173–186.

Cornelius, P. B., B. Gokpinar. 2020. The role of customer investor involvement in crowdfunding success. *Management Science* **66**(1) 452–472.

Du, L., M. Hu, J. Wu. 2017. Contingent stimulus in crowdfunding. *Working Paper, University of Toronto, Toronto.*

Erat, S., S. Kavadias. 2008. Sequential testing of product designs: Implications for learning. *Management Science* **54**(5) 956–968.

Freeman, M., S. Robinson, S. Scholtes. 2021. Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* **67**(7) 4209–4232.

Gambardella, A., C. Raasch, E. von Hippel. 2017. The user innovation paradigm: Impacts on markets and welfare. *Management Science* **63**(5) 1450–1468.

Gao, F., S. Cui, M. Cohen. 2021. Performance, reliability, or time-to-market? Innovative product development and the impact of government regulation. *Production and Operations Management* **30**(1) 253–275.

Ghose, A., P. G. Ipeirotis, B. Li. 2019. Modeling consumer footprints on search engines: An interplay with social media. *Management Science* **65**(3) 1363–1385.

Goodman, J. K., C. Irmak. 2013. Having versus consuming: Failure to estimate usage frequency makes consumers prefer multifeature products. *Journal of Marketing Research* **50**(1) 44–54.

Greene, W. H. 2018. *Econometric Analysis*. Pearson.

Griffiths, T. L., M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* **101** 5228–5235.

Guo, L., J. Zhang. 2012. Consumer deliberation and product line design. *Marketing Science* **31**(6) 995–1007.

Haans, R. F. J., C. Pieters, Z.-L. He. 2016. Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strategic Management Journal* **37**(7) 1177–1195.

Hu, M., X. Li, M. Shi. 2015. Product and pricing decisions in crowdfunding. *Marketing Science* **34**(3) 331–345.

Indiegogo. 2021. Product Stages FAQ- For Backers. http://support.indiegogo.com/hc/en-us/articles/221613968-Product-Stages-FAQ-For-Backers. Accessed on January 2, 2021.

Kesavan, S., B. R. Staats, W. Gilland. 2014. Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science* **60**(8) 1884–1906.

Kickstarter. 2021a. HAIZE - minimalist urban bike navigation. http://www.kickstarter.com/projects/onomo/haize-a-compass-reinvented-navigation-for-urban-cy. Accessed on February 12, 2021.

Kickstarter. 2021b. How can I cancel my project? http://help.kickstarter.com/hc/en-us/articles/115005138393-How-can-I-cancel-my-project-. Accessed on February 22, 2021.

Kickstarter. 2021c. Introducing badges for projects we love. http://www.kickstarter.com/blog/introducing-projects-we-love-badges. Accessed on January 1, 2021.

Kickstarter. 2021d. The Kickstarter blog: Introducing launch now and simplified rules. http://www.kickstarter.com/blog/introducing-launch-now-and-simplified-rules-0/. Accessed on January 1, 2021.

Kickstarter. 2021e. Sphericam 2, the 4k 360° Video Camera for VR. http://www.kickstarter.com/projects/1996234044/sphericam-2-the-4k-360o-video-camera-for-virtual-r/comments. Accessed on May 8, 2021.

Kickstarter. 2021f. Stats. http://www.kickstarter.com/help/stats. Accessed on January 1, 2021.

Kickstarter. 2021g. What do i do if i have questions about a project? http://help.kickstarter.com/hc/en-us/articles/115005126314. Accessed on August 3, 2021.

Koning, R., S. Hasan, A. Chatterji. 2022. Experimentation and start-up performance: Evidence from a/b testing. *Management Science* **68**(9) 6434–6453.

Krishnan, V., K. Ramachandran. 2011. Integrated product architecture and pricing for managing sequential innovation. *Management Science* **57**(11) 2040–2053.

Krishnan, V., K. T. Ulrich. 2001. Product development decisions: A review of the literature. *Management Science* **47**(1) 1–21.

Kuppuswamy, V., B. L. Bayus. 2017. Does my contribution to your crowdfunding project matter? *Journal of Business Venturing* **32**(1) 72–89.

Lévesque, M., M. Minniti, D. Shepherd. 2009. Entrepreneurs' decisions on timing of entry: Learning from participation and from the experiences of others. *Entrepreneurship Theory and Practice* **33**(2) 547–570.

Li, Z., J. A. Duan, S. Ransbotham. 2020. Coordination and dynamic promotion strategies in crowdfunding with network externalities. *Production and Operations Management* **29**(4) 1032–1049.

Lin, M., S. Viswanathan. 2016. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* **62**(5) 1393–1414.

Lin, W., J. M. Wooldridge. 2019. Binary and fractional response models with continuous and binary endogenous explanatory variables. *Working Paper, Jinan University, Guangzhou, China.*

Liu, J., J. Xie, K. K. Yang, Z. Zheng. 2019. Effects of rescheduling on patient no-show behavior in outpatient clinics. *Manufacturing & Service Operations Management* **21**(4) 780–797.

Loch, C., S. Kavadias. 2008. *Handbook of new product development management*. Butterworth-Heinemann, Oxford, UK.

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

Mick, D. G., S. Fournier. 1998. Paradoxes of technology: Consumer cognizance, emotions, and coping strategies. *Journal of Consumer Research* **25**(2) 123–143.

Mollick, E. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* **29**(1) 1–16.

Mollick, E. 2016. The unique value of crowdfunding is not money - it's community. *Harvard Business Review* http://hbr.org/2016/04/the-unique-value-of-crowdfunding-is-not-money-its-community.

Mollick, E., R. Nanda. 2016. Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. *Management Science* **62**(6) 1533–1553.

Özer, Ö., O. Uncu. 2013. Competing on time: An integrated framework to optimize dynamic time-to-market and production decisions. *Production and Operations Management* **22**(3) 473–488.

Petrin, T., K. Train. 2010. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* **47**(1) 3–13.

Roma, P., E. Gal-Or, R. R. Chen. 2018. Reward-based crowdfunding campaigns: Informational value and access to venture capital. *Information Systems Research* **29**(3) 679–697.

Sommer, S. C., C. H. Loch, J. Dong. 2009. Managing complexity and unforeseeable uncertainty in startup companies: An empirical study. *Organization Science* **20**(1) 118–133.

Sorenson, O., V. Assenova, G.-C. Li, J. Boada, L. Fleming. 2016. Expand innovation finance via crowdfunding. *Science* **354**(6319) 1526–1528.

Steyvers, M., T. Griffiths. 2007. Probabilistic topic models. *Handbook of latent semantic analysis* **427**(7) 424–440.

Strausz, R. 2017. A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard. *American Economic Review* **107**(6) 1430–1476.

Tan, T. F., S. Netessine. 2019. When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science* **65**(8) 3495–3517.

Thomke, S., D. E. Bell. 2001. Sequential testing in product development. *Management Science* **47**(2) 308–323.

Thomke, S., D. Reinertsen. 1998. Agile product development: Managing development flexibility in uncertain environments. *California Management Review* **41**(1) 8–30.

Thompson, D. V., R. W. Hamilton, R. T. Rust. 2005. Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research* **42**(4) 431–442.

Tirunillai, S., G. J. Tellis. 2014. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research* **51**(4) 463–479.

Toubia, O., G. Iyengar, R. Bunnell, A. Lemaire. 2019. Extracting features of entertainment products: A guided latent Dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research* **56**(1) 18–36.

Wei, X., M. Fan, W. You, Y. Tan. 2020. An empirical study of the dynamic and differential effects of prefunding. *Production and Operations Management,* Forthcoming.

West, J., M. Bogers. 2014. Leveraging external sources of innovation: A review of research on open innovation. *Journal of Product Innovation Management* **31**(4) 814–831.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Wooldridge, J. M. 2015. Control function methods in applied econometrics. *The Journal of Human Resources* **50**(2) 420–445.

Yoo, O. S., T. Huang, K. Arifoğlu. 2021. A theoretical analysis of the lean start-up method. *Marketing Science,* Forthcoming.

Online Appendix

EC.1. Proofs

In this section, we provide proofs of theoretical results.

Proof of Lemma 1. (a) Suppose that $I \geq 0$, i.e.,

$$\Pi^I = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(2p-2c(q_i+q_u)^2)-C_iq_i-\frac{C_uq_u}{N+1} \geq \Pi^{NI} = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(2p-2cq_i^2)-C_iq_i.$$

This is possible if $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2} \geq \frac{q_i-bq_i^2-p}{q_i-bq_i^2}$, that is $(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2) \geq 0$. So, customer 1 makes comments whenever she pledges, i.e., $v_1 \geq \frac{p}{(q_i+q_u)-b_1(q_i+q_u)^2}$. Thus, $\mathbb{P}(comment) = \frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$ and hence $E[\#of comments] = q_i^n \times \mathbb{P}(comment) = q_i^n\left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)$. When $I \geq 0$, the creator improves the product whenever there are comments, and hence $\mathbb{P}(improve) = \mathbb{P}(comment)$. In this case, customer 2 also pledges with probability $\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}$, and hence $\mathbb{P}(success) = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)^2$.

(b) If $I < 0$, then the creator does not improve the product during the campaign (i.e., $q_f = q_i$) even if customer 1 makes comments, and hence $\mathbb{P}(improve) = 0$. In this case, customer 1 is indifferent between making comments or not, and hence $\mathbb{P}(comment) = \frac{q_i-bq_i^2-p}{q_i-bq_i^2}$ and $E[\#of comments] = q_i^n\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)$ or $\mathbb{P}(comment) = E[\#of comments] = 0$. Also, customer 2 also pledges with probability $\frac{q_i-bq_i^2-p}{q_i-bq_i^2}$, and hence $\mathbb{P}(success) = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)^2$. ■

Proof of Proposition 1. **(a)** From (1), we have

$$\begin{aligned}\frac{\partial I}{\partial q_i} = & -\frac{2p^2(2b)((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2)}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}-\frac{2p^2[(1-2b(q_i+q_u))^2(q_i-bq_i^2)]}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}\\ & -\frac{2p^2((q_i+q_u)-b(q_i+q_u)^2)(1-2bq_i)(1-b(2q_i+q_u))}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}-4c-2c\frac{pb((1-bq_i)+(1-b(q_i+q_u))}{((1-b(q_i+q_u))(1-bq_i))^2}\\ & -nq_i^{n-1}\frac{C_u}{(1+q_i^n)^2}.\end{aligned}$$

Because $1-b(2q_i+q_u) \geq 0$; and by Assumption 1, $1-bq_i > 0$ and $1-b(q_i+q_u) > 0$, $\frac{\partial I}{\partial q_i} < 0$ as long as n is not too small when it is negative. So, we assume that

$$\begin{aligned}nq_i^{n-1}\frac{C_u}{(1+q_i^n)^2} > & -\frac{2p^2(2b)((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2)}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}-\frac{2p^2[(1-2b(q_i+q_u))^2(q_i-bq_i^2)]}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}\\ & -\frac{2p^2((q_i+q_u)-b(q_i+q_u)^2)(1-2bq_i)(1-b(2q_i+q_u))}{(((q_i+q_u)-b(q_i+q_u)^2)(q_i-bq_i^2))^2}-4c-2c\frac{pb((1-bq_i)+(1-b(q_i+q_u))}{((1-b(q_i+q_u))(1-bq_i))^2}\end{aligned}$$

so that $\frac{\partial I}{\partial q_i} < 0$. Also, $\lim_{q_i \to 0^+} I = \infty$ and $\lim_{q_i \to \infty} I = -\infty$. Thus, there exists $\overline{q_i}$ (≥ 0) such that $I \geq 0$ and hence $q_f = q_i + q_u$ if and only if $q_i \leq \overline{q_i}$.

Suppose that $q_i \leq \overline{q_i}$. Then, the first derivative of $\mathbb{P}(comment)$ with respect to q_i is $\frac{\partial \mathbb{P}(comment)}{\partial q_i} = \frac{p(1-2b(q_i+q_u))}{((q_i+q_u)-b(q_i+q_u)^2)^2} > 0$ if and only if $b(q_i+q_u) < 0.5$. Now, suppose that $q_i < \overline{q_i}$. Then, $I < 0$, $q_f =$

q_i, and in one equlibrium, the first derivative of $\mathbb{P}(comment)$ with respect to q_i is $\frac{\partial \mathbb{P}(comment)}{\partial q_i} = \frac{p(1-2bq_i)}{(q_i - bq_i^2)^2} > 0$ if and only if $bq_i < 0.5$. In the other equilibrium, $\mathbb{P}(comment) = 0$.

(b) Suppose that $q_i \leq \overline{q_i}$. Then,

$$\frac{\partial E[\#of comments]}{\partial q_i} = q_i^{n-1}\left(\frac{k((q_i+q_u)-b(q_i+q_u)^2-p)}{(q_i+q_u)-b(q_i+q_u)^2} + \frac{q_i p(1-2b(q_i+q_u))}{((q_i+q_u)-b(q_i+q_u)^2)^2}\right) > 0$$

if and only if

$$n > n' \equiv -\frac{q_i p(1-2b(q_i+q_u))}{((q_i+q_u)-b(q_i+q_u)^2)(((q_i+q_u)-b(q_i+q_u)^2-p))}.$$

Now suppose that $q_i > \overline{q_i}$. Then, in one equilibrium

$$\frac{\partial E[\#of comments]}{\partial q_i} = q_i^{n-1}\left(\frac{n(q_i - bq_i^2 - p)}{q_i - bq_i^2} + \frac{q_i p(1-2bq_i)}{(q_i - bq_i^2)^2}\right) > 0$$

if and only if $n > n'' \equiv -\frac{q_i p(1-2bq_i)}{(q_i - bq_i^2)((q_i - bq_i^2 - p))}$. ■

Proof of Proposition 2. Suppose that $q_i \leq \overline{q_i}$. Then, the first derivative of $\mathbb{P}(improve)$ with respect to q_i is $\frac{\partial \mathbb{P}(improve)}{\partial q_i} = \frac{p(1-2b(q_i+q_u))}{((q_i+q_u)-b(q_i+q_u)^2)^2} > 0$ if and only if $b(q_i + q_u) < 0.5$. Now, suppose that $q_i < \overline{q_i}$. Then, $I < 0$, and hence $\mathbb{P}(comment) = 0$. ■

Proof of Proposition 3. When both $q_f = q_i$ and $q_f = q_i + q_u$, The first derivative of $\mathbb{P}(success)$ with respect to q_i is $\frac{\partial \mathbb{P}(success)}{\partial q_i} = 2\left(\frac{q_i - bq_f^2 - p}{q_f - bq_f^2}\right)\left(\frac{p(1-2bq_f)}{\left(q_f - bq_f^2\right)^2}\right)$. Suppose that $q_i \leq \overline{q_i}$. Then, $I \geq 0$, $q_f = q_i + q_u$, $\frac{\partial \mathbb{P}(success)}{\partial q_i} \geq 0$ if and only if $b(q_i + q_u) \leq 0.5$. Now, suppose that $q_i < \overline{q_i}$. Then, $I < 0$, $q_f = q_i$, and $\frac{\partial \mathbb{P}(success)}{\partial q_i} \geq 0$ if and only if $bq_i < 0.5$. ■

EC.2. Features with Random Value for Customers

In our main analysis, we assume that any additional feature deterministically increases the value that each customer assigns to the product. It is possible that a customer may value some features but not others. In this section, therefore, we consider a case where each customer likes a random fraction of the product features.

In a setting where the number of product features is q_i, we assume that a customer likes $\tilde{q}_i$ features, where $\tilde{q}_i$ follows a Uniform distribution with parameters 0 and q_i. The complexity of the product, though, depends on the actual number of product features q_i. Thus, when the product is not improved during the campaign, each customer i's effective valuation of the product is $v_i \cdot (\tilde{q}_i - bq_i^2)$; and when the product is improved during the campaign, each customer i's effective valuation of the product is $v_i \cdot (\tilde{q}_i + q_u - b(q_i + q_u)^2)$. Here, we assume that there is no uncertainty about q_u as it is suggested by the customer. Keeping the rest of the model as in §2, we numerically analyze these cases according to the setting that we use in §2. Taking the average of randomly generated 10,000 instances, we verify that our theoretical predictions hold.

This analysis enables us to capture the effect of multidimensionality of a design and to show that multidimensionality cannot explain why the probability of campaign success first increases but then decreases. Even if a creator may not be sure whether potential customers will like a certain feature, the creator will believe that an added feature is more likely to be liked than disliked. Thus, the ex-ante probability of campaign success will improve with the number of features as long as each feature is more likely to be liked than disliked.

EC.3. Cost of Commenting

In this section, we consider the case where customer 1 incurs cost of d (>0) when she makes a comment. Suppose that condition (1) holds so that customer 1 anticipates an improvement. Then, customer 1 decides whether to make comments or not by comparing U_1^C when she makes a comment and U_1^{NC} when she does not make a comment, where $U_1^C = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(v_1((q_i+q_u)-b(q_i+q_u)^2)-p)-d$ and $U_1^{NC} = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(v_1(q_i-bq_i^2)-p)$.

Thus, customer 1 makes a comment if and only if $U_1^C \geq U_1^{NC}$, i.e.,

$$v_1 \geq \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}. \tag{EC.1}$$

Suppose that condition (EC.1) holds. Then, in the first stage, customer 1 decides whether to pledge or not by comparing U_1^P when she pledges and U_1^{NP} when she does not pledge, where $U_1^P = U_1^C$ and $U_1^{NP}=0$. Thus, customer 1 pledges if $U_1^P \geq U_1^{NP}$, i.e., $v_1 \geq \frac{p}{(q_i+q_u)-b(q_i+q_u)^2} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}$. Next, suppose that condition (EC.1) does not hold. Then, customer 1 decides whether to pledge or not by comparing $U_1^P = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)(v_1(q_i-bq_i^2)-p)$ and $U_1^{NP}=0$. Thus, customer 1 pledges if and only if $v_1 \geq \frac{p}{q_i-bq_i^2}$. Finally, suppose that condition (1) is violated. Then, customer 1 pledges if and only if $v_1 \geq \frac{p}{q_i-bq_i^2}$. We characterize all possible outcomes of this model in the following lemma.

LEMMA EC.A1. ***(a)*** *Suppose that condition (1) holds.*
(i) *Suppose that* $1 \geq \frac{p}{(q_i+q_u)-b(q_i+q_u)^2} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p} \geq \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}$.

$$\begin{aligned}
&\mathbb{P}(comment) = \mathbb{P}(improve) = \left(1 - \frac{p}{(q_i+q_u)-b(q_i+q_u)^2} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}\right),\\
&E[\#of comments] = q_i^n\left(1 - \frac{p}{(q_i+q_u)-b(q_i+q_u)^2} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}\right),\\
&\mathbb{P}(success) = \left(1 - \frac{p}{(q_i+q_u)-b(q_i+q_u)^2} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}\right)\left(1 - \frac{p}{(q_i+q_u)-b(q_i+q_u)^2}\right).
\end{aligned}$$

(ii) *Suppose that* $\frac{p}{(q_i+q_u)-b(q_i+q_u)^2} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p} < \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)} \leq 1$.

$$\begin{aligned}
&\mathbb{P}(comment) = \mathbb{P}(improve) = \left(1 - \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}\right),\\
&E[\#of comments] = q_i^k\left(1 - \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}\right),\\
&\mathbb{P}(success) = \left(1 - \frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} - \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}\right)\left(1 - \frac{p}{(q_i+q_u)-b(q_i+q_u)^2}\right)\\
&\quad + \left(\frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)} + \frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)} - \frac{p}{q_i-bq_i^2}\right)\left(1 - \frac{p}{q_i-bq_i^2}\right).
\end{aligned}$$

(iii) *Suppose that* $\frac{p}{(q_i+q_u)-b(q_i+q_u)^2}+\frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}\le 1<\frac{p^2}{(q_i-bq_i^2)((q_i+q_u)-b(q_i+q_u)^2)}+\frac{d}{(q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)}$. *Then,* $\mathbb{P}(comment)=\mathbb{P}(improve)=E[\#of comments]=0$ *and* $\mathbb{P}(success)=\left(1-\frac{p}{q_i-bq_i^2}\right)^2$.

(b) *Suppose that condition (1) does not hold. Then,* $\mathbb{P}(comment)=\mathbb{P}(improve)=E[\#of comments]=0$ *and* $\mathbb{P}(success)=\left(1-\frac{p}{q_i-bq_i^2}\right)^2$.

Overall, there are four possible cases. In Case (a), the creator is willing to improve the product further if customer 1 makes a comment. In Case (a-i), customer 1 makes a comment whenever she pledges. (Note that when $\frac{p}{(q_i+q_u)-b(q_i+q_u)^2}+\frac{d}{(q_i+q_u)-b(q_i+q_u)^2-p}>1$, customer 1 never pledges.) In Case (a-ii), customer 1 *may not* make a comment although she pledges. In Case (a-iii), customer 1 *never* makes a comment although she may pledge. In Case (b), the creator is not willing to improve the product further.

We numerically analyze these cases according to the setting that we use in §2 and we select d from Uniform(0,0.1). Taking the average of randomly generated 10,000 instances, we show that our theoretical predictions continue to hold.

EC.4. Additional Benefit of q_u

In this section, we consider the case where q_u can have a different impact than q_i. Specifically, customers can benefit more from features added during the campaign than the product features added before the campaign, and hence the customer's utility can increase more with the addition number of features q_u than the initial number of features q_i. Therefore, when the product is improved during the campaign, each customer i's effective valuation of the product is $v_i\cdot(q_i+a\cdot q_u-b\cdot(q_i+q_u)^2)$, where $a\ge 1$. In this case, $(a-1)q_u$ represents the additional benefit of customer-supported product features to the customer's utility, if any (i.e., $a>1$). This model captures the idea of improving the product during the campaign toward the most desirable paths.

We now explain the differences from our main model in §2. Instead of Assumption 1, we make the following assumption.

Assumption EC.1. *$q_i-bq_i^2>p$ and $(q_i+aq_u)-b(q_i+q_u)^2>p$ such that there exists v_2 such that customer 2's expected utility is positive.*

Following the same steps in backward induction explained in §2, we first revise condition (1) as follows:

$$I\equiv\frac{2p^2(a-b(2q_i+q_u))}{(q_i-bq_i^2)((q_i+aq_u)-b(q_i+q_u)^2)}-2c\left(2q_i+q_u+\frac{p(q_u+(2-a)q_i)}{(1-bq_i)((q_i+aq_u)-b(q_i+q_u)^2)}\right)-C_u\ge 0. \quad \text{(EC.2)}$$

Then, we obtain the following lemma.

LEMMA EC.A2. *(a) Suppose that* $I \geq 0$. *Then* $\mathbb{P}(comment) = \mathbb{P}(improve) = \frac{(q_i+aq_u)-b(q_i+q_u)^2-p}{(q_i+aq_u)-b(q_i+q_u)^2}$, $E[\#of comments] = q_i^n\left(\frac{(q_i+aq_u)-b(q_i+q_u)^2-p}{(q_i+aq_u)-b(q_i+q_u)^2}\right)$, *and* $\mathbb{P}(success) = \left(\frac{(q_i+aq_u)-b(q_i+q_u)^2-p}{(q_i+aq_u)-b(q_i+q_u)^2}\right)^2$.
(b) Suppose that $I < 0$. *Then,* $\mathbb{P}(comment) = \frac{q_i-bq_i^2-p}{q_i-bq_i^2}$ *and* $E[\#of comments] = q_i^n\left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)$ *or* $\mathbb{P}(comment) = E[\#of comments] = 0$. *Also,* $\mathbb{P}(improve) = 0$ *and* $\mathbb{P}(success) = \left(\frac{q_i-bq_i^2-p}{q_i-bq_i^2}\right)^2$.

We numerically analyze these cases according to the setting that we use in §2 and we select a from Uniform(1,2). Taking the average of randomly generated 10,000 instances, we show that our theoretical predictions hold. This analysis enables us to capture the effect of multidimensionality of a design and to show that multidimensionality of a design by itself cannot explain all our empirical observations.

EC.5. Endogenous Addition and Removal of Features during Campaign

In this section, we first theoretically analyze the case where the creator can add or remove an endogenously determined number of product features during the campaign. Then, we empirically test our theoretical prediction about removal of product features and show robustness of our results.

Theoretical Model and Analysis. In this section, we consider the case where customer 1 makes comments to entice the creator to add new feature(s) to the product or remove some of existing feature(s) from the product during the campaign. So, q_u represents the change in the number of features instead of the additional number of features. We only explain the differences from our main model in §2.

As q_u can be negative, the creator's expected profit with an improvement is $\Pi^I = \left(\frac{(q_i+q_u)-b(q_i+q_u)^2-p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(2p-2c(q_i+q_u)^2) - C_iq_i - \frac{C_u}{N+1}|q_u|$. Thus, the creator improves the product during the campaign if and only if $\Pi^I \geq \Pi^{NI}$, i.e.,

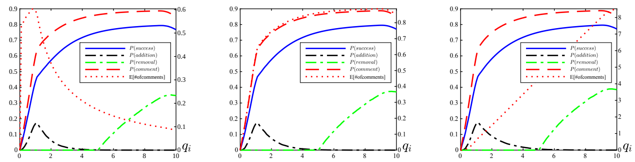
$$I \equiv \frac{2p^2(1-b(2q_i+q_u))q_u}{q_i(q_i+q_u)(1-b(q_i+q_u))(1-bq_i)} - 2cq_u\left(2q_i+q_u-\frac{p}{(1-b(q_i+q_u))(1-bq_i)}\right) - \frac{C_u}{N+1}|q_u| \geq 0. \quad \text{(EC.3)}$$

Suppose that $I \geq 0$. Then, customer 1 makes comments in Stage 2 if and only if $U_1^C \geq U_1^{NC}$, i.e.,

$$v_1((q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)) \geq \frac{p^2\left((q_i+q_u)-b(q_i+q_u)^2-(q_i-bq_i^2)\right)}{(q_i-bq_i^2)(q_i+q_u)-b(q_i+q_u)^2}. \quad \text{(EC.4)}$$

We discuss two possible cases where q_u is endogenously determined. We first consider the case where customer 1 decides on q_u in stage 2 by maximizing her utility. So, the optimal change in the number of features is $q_u^* = arg\ \max_{q_u\in\mathbb{R}}\left(1-\frac{p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(v_1((q_i+q_u)-b(q_i+q_u)^2)-p)$. Evaluating the first-order condition of this utility-maximization problem, we characterize q_u^* as $q_u^* = \frac{1}{2b} - q_i$.

We numerically analyze this case according to the setting that we use in §2 and where $q_u^* = \frac{1}{2b} - q_i$. Taking the average of randomly generated 10,000 instances, we show that our theoretical predictions about $\mathbb{P}(comment)$, $E[\#of comments]$, and $\mathbb{P}(success)$ continue to hold. As it can



(a) $n=-1$. (b) $n=0$. (c) $n=1$.

Figure EC.5.1 The impact of the initial development level q_i on $\mathbb{P}(comment)$, $E[\#of comments]$, $\mathbb{P}(improve)$, and $\mathbb{P}(success)$. The average of randomly generated 10,000 instances. The setting is the same as in Figure 3.

be seen from Figure EC.5.1, we also show that as q_i increases, the probability that the creator adds feature(s) during the campaign (i.e., $\mathbb{P}(addition)$) first increases but then decreases; and the probability that the creator removes feature(s) during the campaign (i.e., $\mathbb{P}(removal)$) first increases but then decreases. The intuition is as follows. When q_i is small, the customer prefers an increase in the number of features, and hence q_u is positive. In this case, $\mathbb{P}(addition)$ first increases because the probability that the customer pledges and makes comments increases, but then $\mathbb{P}(addition)$ decreases because it becomes too costly for the creator to make any additions. On the other hand, when q_i is large, the customer prefers a decrease in the number of features, and hence q_u is negative. Again in this case, $\mathbb{P}(removal)$ first increases but then decreases. Notice that when q_i is moderate, $\mathbb{P}(addition)$ and $\mathbb{P}(removal)$ are very small because in this case, q_i is very close to the number of features that maximizes the creator's profit. Thus, the additional cost of adding or removing a feature can not be compensated by the small increase in the chance of campaign success.

In addition to the case where customer 1 decides on q_u in stage 2, we also show similar results when the creator decides on q_u in stage 3, and hence the optimal change in the number of features is $q_u^* = arg\ \max_{q_u\in\mathbb{R}}\left(1-\frac{p}{(q_i+q_u)-b(q_i+q_u)^2}\right)(2p-2c(q_i+q_u)^2)-C_iq_i-\frac{C_u}{N+1}|q_u|$.

This analysis provides two key findings. First, our main result about the impact of q_i on $\mathbb{P}(improve)$ for any $q_u(>0)$ in §2.3 continues to hold when q_u is optimized either by the customer or by the creator and when q_u^* is positive. Second, when q_i is large, the likelihood of removal of features from the product first increases and then decreases.

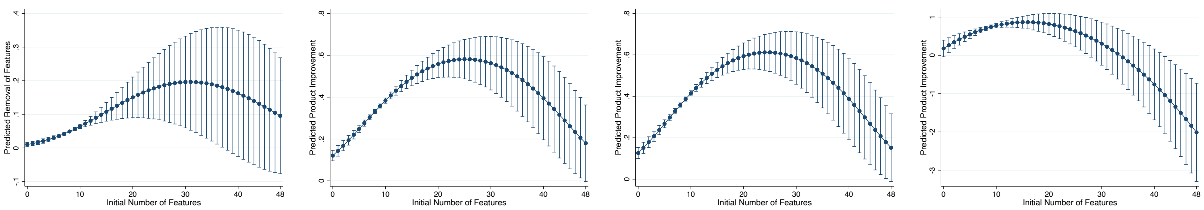
Next, we empirically test the impact of q_i on $\mathbb{P}(removal)$, and also show the robustness of our results considering the alternative definitions of product improvement.

Empirical Model and Analysis. To test the impact of the initial number of features on removal of features, we define removal of features R_k for each campaign k. Specifically, $R_k = 1$ if $q_{fk} < q_{ik}$; otherwise, $R_k = 0$. We observe that in 6% of campaigns, $q_{fk} < q_{ik}$. Replacing I_k in our IV Model 2 in §3.3 with R_k, we obtain the results in column (1) of Table EC.5.1. As Figure EC.5.2(a) illustrates, the likelihood of removal of features first increases and then decreases with the initial number of

Table EC.5.1 Results of IV model 2 with alternative definitions of product improvement.

	(1) Second Stage of IV Model 2 Removal of features	(2) Second Stage of IV Model 2 Removal of features	(3) Second Stage of IV Model 2 Product improvement (Alternative definition 1)	(4) Second Stage of IV Model 2 Product improvement (Alternative definition 2)	(5) Second Stage of IV Model 2 Product improvement (Alternative definition 3)
Initial number of features	.071***	.097***	.108***	.114***	.102***
	(.013)	(.014)	(.007)	(.007)	(.014)
Initial number of features²	-.002***	-.002***	-.002***	-.002***	-.002***
	(0)	(0)	(0)	(0)	(0)
Residuals	.013	.007	-.023***	-.021***	-.06***
	(.013)	(.015)	(.007)	(.007)	(.012)
Residuals²	0	0	0	0	0
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-2.469***	-2.464***	-1.122***	-1.152***	.2*
	(.154)	(.136)	(.083)	(.081)	(.12)
Wald χ^2	1288.29	1270.06	2562.33	2709.80	319.41
R^2 or Pseudo R^2	.089	.134	.089	.105	.018
Observations	18,173	13,456	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*



(a) Removal of Features (b) Alternative Definition 1 (c) Alternative Definition 2 (d) Alternative Definition 3

Figure EC.5.2 Predicted likelihood of removal of features and product improvement.

features, supporting our theoretical prediction. We also run the same model after we remove 4,717 observations where $q_{fk} > q_{ik}$. As column (2) of Table EC.5.1 shows, our results continue to hold.

Robustness Analyses about Removals. As an additional robustness check, we also include both removal and addition of product features while defining product improvement I_k for each campaign k. Specifically, $I_k = 1$ if $q_{fk} > q_{ik}$ or $q_{fk} < q_{ik}$; otherwise, $I_k = 0$. We observe that in 26% of campaigns, $q_{fk} > q_{ik}$; and in 6% of campaigns, $q_{fk} < q_{ik}$. As shown in column (3) of Table EC.5.1, our results about H2 continue to hold when we consider the decrease in the number of product features as product improvement (see Figures EC.5.2(b)).

Additionally, we use another alternative definition of product improvement. Specifically, instead of comparing the final number of product features with the initial number of product features, we analyze any addition or removal of a feature during the campaign. By this way, we identify additional 394 campaigns where $q_{fk} = q_{ik}$ but there is a change in the existing features during the campaign, and we classify them as $I_k = 1$ in addition to campaigns where $q_{fk} > q_{ik}$ or $q_{fk} < q_{ik}$. As shown in column (4) of Table EC.5.1, our results about H2 continue to hold (see Figures EC.5.2(c)). Finally, as shown in column (5) of Table EC.5.1, our results about H2 continue to hold when we define a continuous measure of product improvement, i.e., $I_k = q_{fk} - q_{ik}$ (see Figures EC.5.2(d)).

EC.6. Impact of Existence of Comments on Product Improvement

In this section, we analyze the impact of the existence of comments EC_k on product improvement I_k. As in all IV models, we regress the initial number of features q_{ik} on the instrumental variable B_k and control variables in the first stage, and obtain the predicted residuals $\hat{u}_k$ to use in the second stage. Since our aim is to analyze the impact of the existence of comments EC_k on product improvement I_k, we have two steps in the second stage. First, as in our IV Model 1a, we analyze the exogenous impact of q_{ik} on the existence of comments EC_k. Second, as in our IV Model 2, we analyze the the exogenous impact of q_{ik} on product improvement I_k, but this time we add the existence of comments EC_k to this regression. Therefore, we obtain the following first-stage regression and two second-stage regressions, respectively:

$$q_{ik} = \alpha_0 + \alpha_1 B_k + \alpha_X X_k + u_k,$$

$$P(EC_k) = \Phi(\beta_0 + \beta_1 q_{ik} + \beta_2 (q_{ik})^2 + \beta_3 \hat{u}_k + \beta_4 (\hat{u}_k)^2 + \beta_X X_k + v_k), \text{ and}$$

$$P(I_k) = \Phi(\gamma_0 + \gamma_1 q_{ik} + \gamma_2 (q_{ik})^2 + \gamma_3 EC_k + \gamma_4 \hat{u}_k + \gamma_5 (\hat{u}_k)^2 + \gamma_X X_k + z_k).$$

As our two dependent variables in the second stage are binary, we use a biprobit model to jointly

Table EC.6.1 Impact of Customer Feedback on Product Improvement

	(1) First Stage of IV Models Initial number of features	(2) Second Stage of IV Model with Biprobit Existence of comment(s)	(3) Second Stage of IV Model with Biprobit Product improvement	(4) Second Stage of IV Model with Biprobit Existence of comment(s)	(5) Second Stage of IV Model with Biprobit Product improvement
Before relaxation of rules	3.051***				
	(.168)				
Customers' previous pledges	-.001***	0***	-.001***	-.001***	-.001***
	(0)	(0)	(0)	(0)	(0)
Customers' previous comments (log)				5.106***	
				(.438)	
Initial number of features		.134***	.052***	.141***	.056***
		(.012)	(.016)	(.013)	(.013)
Initial number of features²		-.002***	-.002***	-.002***	-.002***
		(0)	(0)	(0)	(0)
Residuals		-.064***	-.002	-.071***	-.004
		(.011)	(.012)	(.012)	(.012)
Residuals²		.001**	0	.001**	0
		(0)	(0)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-2.025***	-.9***	-1.195***	-1.116***	-1.184***
	(.465)	(.093)	(.1)	(.095)	(.099)
Existence of comment(s)			.834***		.746***
			(.169)		(.062)
ρ			-.189***		-.145***
			(.104)		(.036)
Wald χ^2	7120.96	5589.41	5589.41	5896.59	5896.59
R^2 or Pseudo R^2	.415				
Observations	13,568	13,568	13,568	13,568	13,568

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

estimate them (e.g., Liu et al. 2019, Freeman et al. 2021). Columns (1), (2), and (3) of Table EC.6.1 show the results of the first-stage regression and two steps in the second-stage (i.e., biprobit). As it can be seen from columns (2) and (3) of Table EC.6.1, the impact of the initial number of features on the likelihood of the existence of comments ($\beta_1 = 0.134$ and $\beta_2 = -0.002$, $p < 0.01$) and on the likelihood of product improvement ($\gamma_1 = 0.052$ and $\gamma_2 = -0.002$, $p < 0.01$) continue to hold. Also, as it can be seen from column (3) of Table EC.6.1, the coefficient of the existence of comments EC_k is positive and significant ($\gamma_3 = 0.834$, $p < 0.01$), which suggests that the likelihood of product improvement increases with the existence of comments. Although our biprobit model can be estimated without an instrumental variable for the existence of comments EC_k (e.g., Freeman et al. 2021), we use $log_{10}(Customers\text{'}\ previous\ comments + 1)$ as the instrumental variable for the existence of comments EC_k. As columns (4) and (5) of Table EC.6.1 show, the results are very similar to the results of the model without the instrumental variable.

Overall, these analyses show that our main results about the impact of the initial number of features on the existence of comments and product improvement continue to hold when we consider

Table EC.7.1 Mediating role of product improvement

	(1) First Stage of IV Model	(2) Second Stage of IV Model	(3) First Stage of IV Model	(4) Second Stage of IV Model	(5) First Stage of IV Model	(6) Second Stage of IV Model
	Number of improvements	Campaign Success	Number of improvements	Campaign Success	Number of improvements	Campaign Success
Before relaxation of rules	.168***		.1*			
	(.04)		(.057)			
Customers' previous pledges					-.001***	-.007***
					(0)	(0)
Customers' previous comments (log)					1.274***	
					(.232)	
Number of improvements		1.494***		2.278		1.134***
		(.525)		(215.368)		(.201)
Number of improvements ^{2}		.006		.004		.014
		(.018)		(.022)		(.018)
Residuals		-1.398***		-2.167		-1.042***
		(.527)		(215.369)		(.202)
Residuals ^{2}		-.005		-.003		-.013
		(.018)		(.022)		(.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	.253**	1.768***	.389**	1.269	.517***	2.289***
	(.124)	(.3)	(.168)	(130.138)	(.145)	(.214)
Wald χ^2	150.49	1576.92	127.60	472.63	191.63	1438.27
R^2 or Pseudo R^2	0.007	.206	0.009	.195	0.014	.254
Observations	18,173	18,173	13,568	13,568	13,568	13,568

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

the impact of the existence of comments on product improvement; and the existence of comments has a positive effect on the likelihood of product improvement.

EC.7. Mediation Analysis

Impact of Product Improvements. In §3.6, we analyze the exogenous impact of product improvements. In this section, we analyze the mediating role of product improvements.

We first analyze the impact of the *initial number of features* (q_{ik}) on *campaign success* (S_k) through product improvements. In this analysis, product improvement is the mediator, and as we use the control function approach, we generate a continuous measure which is the *number of improvements* $NI_k = q_{fk} - q_{ik}$ for each campaign k. As our instrumental variable for q_{ik}—*before relaxation of rules* (B_k)—does not directly affect product improvements, we can use it to tease out the mediation. Specifically, by instrumenting the *number of improvements* NI_k with B_k, we estimate the impact the impact of q_{ik} on S_k mediated through NI_k. We obtain the following IV model (IV Model M1) with the first-stage and second-stage regressions, respectively:

$$
\begin{aligned}
NI_k &= \alpha_0 + \alpha_1 B_k + \alpha_X X_k + u_k, \text{ and} \\
P(S_k) &= \Phi(\beta_0 + \beta_1 \cdot NI_k + \beta_2 (NI_k)^2 + \beta_3 \hat{u}_k + \beta_4 (\hat{u}_k)^2 + \beta_X X_k + v_k).
\end{aligned}
$$

We next analyze the impact of the number of comments on campaign success through product improvements. As the instrumental variable, *customers' previous comments* (PC_k), does not have a direct effect on product improvements (NI_k), we use it to tease out the mediation following the same approach as above. (Note that in this case, we control for the average number of previous campaigns that those customers pledge, i.e., *customers' previous pledges.*) So, we obtain the following IV model (IV Model M2) with the first-stage and second-stage regressions, respectively:

$$
\begin{aligned}
NI_k &= \gamma_0 + \gamma_1 PC_k + \gamma_X X_k + u_k, \text{ and} \\
P(S_k) &= \Phi(\theta_0 + \theta_1 \cdot NI_k + \theta_2 (NI_k)^2 + \theta_3 \hat{u}_k + \theta_4 (\hat{u}_k)^2 + \theta_X X_k + v_k).
\end{aligned}
$$

Table EC.7.1 summarizes the results of the mediation analysis. Column (2) of Table EC.7.1 shows that the likelihood of campaign success increases with product improvements as a result of the

Table EC.7.2 Mediating role of customer feedback

	(1) First Stage of IV Model Number of comments	(2) Second Stage of IV Model Product improvement
Before relaxation of rules	2.016***	
	(.375)	
Number of comments		.096***
		(.017)
Number of comments^2		0***
		(0)
Residuals		-.069***
		(.017)
Residuals^2		0***
		(0)
Controls	Yes	Yes
Constant	-17.461***	.413
	(1.957)	(.301)
Wald χ^2	1109.71	573.93
R^2 or Pseudo R^2	.072	.064
Observations	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

initial number of features ($\beta_1 = 1.494$, $p < 0.01$; and $\beta_2 = 0.006$, $p > 0.1$). Similarly, column (6) of Table EC.7.1 shows that the likelihood of campaign success increases with product improvements as a result of the number of comments ($\theta_1 = 1.134$, $p < 0.01$; and $\theta_2 = 0.014$, $p > 0.1$), consistent with Cornelius and Gokpinar (2020).

Recall that the data set about customers' previous comments is available for 13,568 campaigns launched before August 2015. When we run IV Model M1 using this smaller data set, as column (4) of Table EC.7.1 shows, the impact of product improvements is weaker.

Impact of Customer Feedback. In §3.6, we analyze the exogenous impact of customer feedback. In this section, we analyze the mediating role of customer feedback. Specifically, we first analyze the impact of the *initial number of features* (q_{ik}) on *product improvement* (I_k) through *number of comments* (NC_k). In this analysis, the *number of comments* (NC_k) is the mediator. As our instrumental variable for q_{ik}—*before relaxation of rules* (B_k)—does not directly affect the number of comments, we can use it to tease out the mediation. Specifically, by instrumenting the *number of comments* NC_k with B_k, we estimate the impact the impact of q_{ik} on I_k mediated through NC_k. We obtain the following IV model (IV Model M3) with the first-stage and second-stage regressions, respectively:

$$
\begin{aligned}
NC_k &= \alpha_0 + \alpha_1 B_k + \alpha_X X_k + u_k, \text{ and} \\
P(I_k) &= \Phi(\beta_0 + \beta_1 \cdot NC_k + \beta_2 (NC_k)^2 + \beta_3 \hat{u}_k + \beta_4 (\hat{u}_k)^2 + \beta_X X_k + v_k).
\end{aligned}
$$

Table EC.7.2 summarizes the results of the mediation analysis. Column (2) of Table EC.7.2 shows that the likelihood of product improvement increases with the number of comments as a result of the initial number of features ($\beta_1 = 0.096$ and $\beta_2 = 0$, $p < 0.01$).

EC.8. Details of LDA Model

In this section, we discuss the details of the LDA method (Blei et al. 2003). The LDA method assumes that each document can be represented as a mixture of topics and each topic can be represented as a mixture of words. So, taking a corpus of documents as an input, the LDA method outputs the distribution of topics in each document and the distribution of words in each topic. The distribution of topics in each document is a vector of weights, where the weight of each topic

HAIZE is a new type of navigation system designed for urban cyclists. It works like a magic compass that, instead of pointing north, points to the destination you set in our app. HAIZE leaves you free to choose your own route through the city. It also makes your ride safer by letting you keep your phone in your pocket. HAIZE can easily be attached to any bike and is small and built-to-last so that you can always bring it with you. Our companion app will be available for both Android- and iOS-based smartphones. HAIZE is easy to use. Simply attach it to your bike, set the destination in our companion app, put your phone away and let HAIZE guide you. Its simple LED-based display will point you in the right direction and let you know the distance to your destination. You can select between two different modes of navigation. The compass mode points you in the direction of your final destination and lets you explore along the way. The navigation mode sets a specific route and gives you turn-by-turn directions. Check out the video of HAIZE in action: Check out this video of HAIZE and our app: HAIZE is stripped down to the essence, both the led-based display and the aluminium body combine simplicity with usability. The HAIZE led-display gives you all the information you need at a glance; direction and distance. HAIZE lets you focus on the road and explore the city. If you are interested in additional details about your trip, you can always check the app after your ride. Here you will find stats about your trips and saved routes. The rubber band integrated in the HAIZE body allows you to easily attach it on any bike and keep HAIZE comfortably in your pocket when leaving your bike on the street. The body of HAIZE has been created using machined anodized aluminium, making it both sturdy and stylish. The magnetometer tracks the direction to the destination The accelerometer and gyroscope are used to determine HAIZEs position The light sensor is used to regulate the LED brightness to accommodate different lighting conditions The battery with 300mAh delivers 2 weeks of normal usage and can be easily recharged using a micro usb connection HAIZE is connected to our app via a low power bluetooth 4.0 connection As you can see there is a lot of technology packed into HAIZE...

Existing Topic	0.058*"control" + 0.052*"smart" + 0.037*"sensor" + 0.030*"home" + 0.021*"mode" + 0.021*"button" + 0.019*"connect" + 0.017*"set" + 0.017*"remot" + 0.017*"monitor"

(a) Excerpt from the initial description where words corresponding to an existing feature are highlighted.

HAIZE is minimalist navigation device for urban cyclists. It is designed focusing on high quality materials, style and simplicity. HAIZE works like a magic compass that, instead of pointing north, points to the destination you set in our app. HAIZE leaves you free to choose your own route through the city. It also makes your ride safer by letting you keep your phone in your pocket. If you feel like sticking to the main roads HAIZE also offers turn-by-turn navigation. WHY YOU NEED ONE? 2 navigation modes: "turn-by-turn mode" and "compass mode" Self-regulating LED display for perfect day and night-time visibility Sturdy and high-quality materials Can be used on any bike Wristband to use HAIZE while running, hiking or geo-caching HAIZE is easy to use. Simply attach it to your bike, set the destination in our companion app, put your phone away and let HAIZE guide you. Its simple LED-based display will point you in the right direction and let you know the distance to your destination. You can select between two different modes of navigation. The compass mode points you in the direction of your final destination and lets you explore along the way. The navigation mode sets a specific route and gives you turn-by-turn directions. Check out the video of HAIZE in action: Check out this video of HAIZE and our app: HAIZE is stripped down to the essence, both the led-based display and the aluminium body combine simplicity with usability. HAIZE is made out of aeronautic-grade sandblasted aluminium and shockproof glass. A tested and durable combination that stands out from the first moment. To know a direction you don\'t need to get distracted processing numbers or symbols on a screen. The HAIZE led-display gives you all the information you need at a glance; direction and distance in a simple and intuitive way. HAIZE lets you focus on the road and explore the city. If you are interested in additional details about your trip, you can check them in the app. There you will find stats about your trips and saved routes. HAIZE automatically regulates the brightness of the LEDs to work perfectly under any light condition. It will help you navigate the city no matter what time of the day! The elastic band integrated in the HAIZE body allows you to easily attach it on any bike and keep HAIZE comfortably in your pocket when leaving your bike on the street. And it always stays in place! HAIZE was originally designed for urban cycling. But many of our backers wanted to use it in other situations. That is why we decided to give every backer a wristband to bring HAIZE along to any activity. Be it for hiking, running, or geo-caching. And of course finding your way back to last years perfect mushroom spot. HAIZE will be able to guide you to the best spots while wandering freely. And you can be confident about getting back to the basecamp no matter how many turns you make. The HAIZE wristband is made from high quality silicone and fits perfectly around the aluminum case, allowing you to take HAIZE everywhere. The magnetometer tracks the direction to the destination The accelerometer and gyroscope are used to determine HAIZEs position The light sensor is used to regulate the LED brightness to accommodate different lighting conditions The battery with 300mAh delivers 2 weeks of normal usage and can be easily recharged using a micro usb connection HAIZE is connected to our app via a low power bluetooth 4.0 connection As you can see there is a lot of technology packed into HAIZE...

Existing Topic	0.058*"control" + 0.052*"smart" + 0.037*"sensor" + 0.030*"home" + 0.021*"mode" + 0.021*"button" + 0.019*"connect" + 0.017*"set" + 0.017*"remot" + 0.017*"monitor"
Added Topic	0.059*"materi" + 0.038*"weight" + 0.032*"high" + 0.030*"durabl" + 0.027*"surfac" + 0.023*"made" + 0.022*"blade" + 0.021*"strong" + 0.020*"strength" + 0.020*"resist"

(b) Excerpt from the final description where words corresponding to an existing feature and an added feature are highlighted.

Figure EC.8.1 Initial and final descriptions of the product HAIZE with examples of an "existing" topic that is available in the initial description and an "added" topic that is added to the final description. Tables below excerpts illustrate the most relevant ten words with their weights in these topics.

represents how intensively the topic is used in the document. Similarly, the distribution of words in each topic represents the frequency of words. As product descriptions on campaign pages include explanation of product features, the LDA method is suitable for extracting topics related to these features from product descriptions (e.g., Tirunillai and Tellis 2014, Toubia et al. 2019).

To train the LDA model, we start with 43,536 initial and final descriptions of products in 21,768 campaigns. Following the standard practice (e.g., Tirunillai and Tellis 2014, Toubia et al. 2019), we first pre-process descriptions (e.g., remove stop words, remove descriptions that contain less than ten words and that are not written in English, and stem words). We then fit the LDA model on the corpus of the remaining 42,564 descriptions (from 21,380 campaigns, 196 of which only have a single description after pre-processing) using the standard hyperparameters of $\alpha = 1$ and $\beta = 0.01$ (e.g., Steyvers and Griffiths 2007, Toubia et al. 2019, Ghose et al. 2019), where α and β are parameters of the prior Dirichlet distributions of topics in documents and words in topics, respectively (Blei et al. 2003). Following the rule of $\alpha = 50/T$, where T is the number of topics, (e.g., Steyvers and Griffiths 2007, Tirunillai and Tellis 2014), we set the number of topics to 50. From the trained LDA model, we obtain weights of words in each of 50 topics and weights of topics in each of 42,564 descriptions; all weights are positive. See Table EC.8.1 for all topics with most frequent words and Figure EC.8.1 for an example product description and corresponding topics.

Table EC.8.1 Topics with most frequent words

Topics	Number of Campaigns	Words in Topics with Respective Weights
0	8,157	0.108*"video" + 0.041*"display" + 0.040*"see" + 0.039*"imag" + 0.036*"screen" + 0.033*"view" + 0.031*"photo" + 0.030*"pictur" + 0.029*"show" + 0.019*"digit"
1	15,651	0.098*"prototyp" + 0.070*"design" + 0.069*"product" + 0.057*"final" + 0.057*"test" + 0.043*"first" + 0.025*"readi" + 0.022*"work" + 0.022*"complet" + 0.019*"concept"
2	5,969	0.058*"car" + 0.052*"energi" + 0.040*"electr" + 0.039*"generat" + 0.031*"vehicl" + 0.028*"use" + 0.027*"save" + 0.020*"effici" + 0.020*"cost" + 0.019*"wind"
3	7,202	0.066*"board" + 0.038*"use" + 0.029*"modul" + 0.024*"sourc" + 0.024*"code" + 0.022*"open" + 0.021*"hardwar" + 0.020*"control" + 0.019*"program" + 0.017*"project"
4	8,925	0.037*"day" + 0.031*"page" + 0.027*"week" + 0.027*"month" + 0.025*"word" + 0.024*"goal" + 0.022*"note" + 0.022*"read" + 0.022*"paper" + 0.021*"share"
5	3,478	0.114*"card" + 0.067*"wallet" + 0.061*"pocket" + 0.038*"block" + 0.032*"slim" + 0.025*"carri" + 0.022*"credit" + 0.019*"slot" + 0.016*"hold" + 0.016*"back"
6	9,968	0.231*"one" + 0.117*"two" + 0.076*"small" + 0.054*"line" + 0.046*"three" + 0.033*"larg" + 0.027*"size" + 0.024*"four" + 0.018*"differ" + 0.018*"first"
7	4,949	0.139*"water" + 0.056*"air" + 0.027*"plant" + 0.024*"use" + 0.021*"grow" + 0.017*"pressur" + 0.016*"natur" + 0.015*"environ" + 0.015*"garden" + 0.014*"shower"
8	7,418	0.057*"patent" + 0.045*"comfort" + 0.038*"problem" + 0.030*"bodi" + 0.029*"posit" + 0.028*"head" + 0.027*"sleep" + 0.024*"help" + 0.023*"solut" + 0.022*"invent"
9	8,066	0.113*"part" + 0.105*"build" + 0.065*"kit" + 0.055*"assembl" + 0.031*"step" + 0.030*"need" + 0.028*"make" + 0.026*"built" + 0.024*"includ" + 0.021*"set"
10	9,775	0.068*"easi" + 0.046*"fit" + 0.041*"easili" + 0.034*"use" + 0.034*"quick" + 0.034*"remov" + 0.028*"attach" + 0.026*"simpl" + 0.021*"simpli" + 0.020*"clip"
11	12,480	0.060*"team" + 0.053*"develop" + 0.049*"technolog" + 0.038*"year" + 0.034*"engin" + 0.029*"world" + 0.024*"experi" + 0.024*"innov" + 0.021*"industri" + 0.016*"research"
12	18,106	0.050*"would" + 0.035*"like" + 0.031*"could" + 0.022*"time" + 0.021*"idea" + 0.020*"look" + 0.018*"work" + 0.017*"tri" + 0.016*"start" + 0.015*"think"
13	8,242	0.065*"unit" + 0.040*"current" + 0.036*"measur" + 0.034*"compon" + 0.031*"use" + 0.027*"test" + 0.022*"suppli" + 0.021*"requir" + 0.019*"standard" + 0.018*"wire"
14	4,482	0.061*"heat" + 0.052*"bottl" + 0.044*"glass" + 0.036*"pour" + 0.033*"coffe" + 0.031*"hot" + 0.030*"temperatur" + 0.029*"cup" + 0.025*"cold" + 0.024*"drink"
15	6,209	0.055*"steel" + 0.053*"machin" + 0.045*"pen" + 0.033*"finish" + 0.032*"metal" + 0.032*"stainless" + 0.030*"aluminum" + 0.025*"plate" + 0.023*"cap" + 0.022*"titanium"
16	5,178	0.096*"bag" + 0.056*"leather" + 0.043*"travel" + 0.037*"strap" + 0.028*"fabric" + 0.026*"carri" + 0.025*"pack" + 0.025*"belt" + 0.023*"pocket" + 0.017*"cloth"
17	6,097	0.087*"box" + 0.043*"shirt" + 0.041*"edit" + 0.034*"limit" + 0.029*"seri" + 0.029*"set" + 0.026*"figur" + 0.020*"name" + 0.019*"anim" + 0.019*"origin"
18	13,177	0.076*"reward" + 0.070*"ship" + 0.063*"campaign" + 0.049*"backer" + 0.049*"pledg" + 0.039*"goal" + 0.031*"receiv" + 0.031*"add" + 0.025*"stretch" + 0.023*"pleas"
19	10,970	0.073*"creat" + 0.039*"support" + 0.037*"communiti" + 0.030*"help" + 0.028*"creativ" + 0.023*"art" + 0.023*"dream" + 0.021*"world" + 0.020*"work" + 0.017*"share"
20	13,771	0.208*"product" + 0.056*"market" + 0.041*"compani" + 0.030*"cost" + 0.028*"custom" + 0.027*"price" + 0.025*"manufactur" + 0.025*"need" + 0.025*"order" + 0.023*"busi"
21	16,688	0.062*"want" + 0.057*"make" + 0.053*"get" + 0.032*"know" + 0.027*"peopl" + 0.023*"need" + 0.022*"help" + 0.020*"thing" + 0.019*"let" + 0.019*"way"
22	10,828	0.032*"provid" + 0.029*"peopl" + 0.027*"person" + 0.026*"inform" + 0.019*"servic" + 0.016*"individu" + 0.016*"user" + 0.014*"busi" + 0.014*"site" + 0.014*"locat"
23	3,626	0.076*"sound" + 0.062*"music" + 0.049*"record" + 0.036*"audio" + 0.033*"speaker" + 0.031*"play" + 0.021*"qualiti" + 0.019*"headphon" + 0.018*"listen" + 0.017*"hear"
24	6,136	0.063*"camera" + 0.062*"mount" + 0.048*"mold" + 0.043*"lock" + 0.028*"use" + 0.027*"arm" + 0.026*"angl" + 0.021*"holder" + 0.019*"inject" + 0.018*"design"
25	7,303	0.137*"color" + 0.064*"black" + 0.046*"red" + 0.045*"option" + 0.043*"choos" + 0.042*"avail" + 0.036*"choic" + 0.034*"white" + 0.034*"blue" + 0.028*"colour"
26	4,220	0.214*"light" + 0.082*"led" + 0.031*"use" + 0.031*"night" + 0.025*"len" + 0.022*"bright" + 0.021*"switch" + 0.017*"turn" + 0.016*"lens" + 0.016*"lamp"
27	7,599	0.058*"control" + 0.052*"smart" + 0.037*"sensor" + 0.030*"home" + 0.021*"mode" + 0.021*"button" + 0.019*"connect" + 0.017*"set" + 0.017*"remot" + 0.017*"monitor"
28	9,342	0.059*"materi" + 0.038*"weight" + 0.032*"high" + 0.030*"durabl" + 0.027*"surfac" + 0.023*"made" + 0.022*"blade" + 0.021*"strong" + 0.020*"strength" + 0.020*"resist"
29	6,324	0.056*"data" + 0.039*"access" + 0.032*"secur" + 0.030*"use" + 0.026*"web" + 0.025*"dog" + 0.025*"network" + 0.021*"user" + 0.021*"storag" + 0.020*"comput"
30	7,547	0.057*"made" + 0.057*"piec" + 0.048*"wood" + 0.041*"natur" + 0.038*"beauti" + 0.035*"make" + 0.028*"materi" + 0.027*"shape" + 0.023*"uniqu" + 0.022*"tree"
31	5,947	0.135*"case" + 0.108*"phone" + 0.079*"protect" + 0.052*"cover" + 0.021*"use" + 0.017*"anti" + 0.016*"also" + 0.014*"call" + 0.014*"safe" + 0.014*"back"
32	8,568	0.203*"project" + 0.061*"fund" + 0.034*"rais" + 0.029*"goal" + 0.023*"help" + 0.021*"hope" + 0.021*"money" + 0.021*"flight" + 0.020*"fli" + 0.018*"need"
33	11,045	0.220*"design" + 0.053*"qualiti" + 0.027*"style" + 0.025*"uniqu" + 0.023*"high" + 0.018*"look" + 0.018*"perfect" + 0.017*"function" + 0.016*"collect" + 0.016*"classic"
34	3,889	0.088*"bike" + 0.045*"frame" + 0.042*"ride" + 0.034*"road" + 0.032*"tube" + 0.026*"bicycl" + 0.025*"tire" + 0.025*"front" + 0.023*"seat" + 0.023*"cycl"
35	10,185	0.155*"new" + 0.054*"full" + 0.047*"featur" + 0.027*"plus" + 0.025*"rang" + 0.021*"includ" + 0.019*"great" + 0.019*"first" + 0.018*"offer" + 0.016*"best"
36	13,723	0.027*"use" + 0.022*"may" + 0.022*"mani" + 0.021*"howev" + 0.020*"differ" + 0.017*"result" + 0.016*"effect" + 0.015*"form" + 0.014*"requir" + 0.013*"possibl"
37	7,359	0.057*"famili" + 0.056*"friend" + 0.047*"love" + 0.039*"fun" + 0.037*"children" + 0.023*"child" + 0.022*"help" + 0.021*"toy" + 0.020*"stori" + 0.019*"play"
38	5,796	0.042*"wheel" + 0.026*"gear" + 0.023*"bar" + 0.023*"roll" + 0.020*"get" + 0.020*"ski" + 0.017*"run" + 0.017*"mountain" + 0.016*"feet" + 0.016*"race"
39	4,949	0.061*"clean" + 0.042*"food" + 0.034*"skin" + 0.034*"use" + 0.025*"organ" + 0.022*"contain" + 0.021*"safe" + 0.020*"make" + 0.019*"dri" + 0.019*"time"
40	6,710	0.159*"devic" + 0.064*"cabl" + 0.044*"mobil" + 0.037*"tablet" + 0.034*"use" + 0.028*"appl" + 0.025*"adapt" + 0.025*"work" + 0.021*"support" + 0.020*"station"
41	6,830	0.078*"top" + 0.068*"space" + 0.037*"side" + 0.034*"design" + 0.033*"place" + 0.032*"room" + 0.028*"home" + 0.027*"tabl" + 0.026*"ball" + 0.024*"sit"
42	4,246	0.133*"watch" + 0.033*"time" + 0.032*"movement" + 0.029*"fit" + 0.027*"train" + 0.025*"sport" + 0.025*"band" + 0.020*"strap" + 0.020*"exercis" + 0.018*"activ"
43	8,330	0.117*"system" + 0.038*"perform" + 0.028*"motor" + 0.028*"speed" + 0.025*"engin" + 0.020*"control" + 0.018*"high" + 0.016*"forc" + 0.016*"provid" + 0.015*"mechan"
44	5,661	0.156*"power" + 0.110*"charg" + 0.098*"batteri" + 0.025*"charger" + 0.021*"plug" + 0.020*"time" + 0.020*"panel" + 0.018*"hour" + 0.017*"cell" + 0.014*"portabl"
45	5,660	0.082*"key" + 0.074*"open" + 0.055*"ring" + 0.050*"magnet" + 0.047*"inch" + 0.039*"wall" + 0.033*"stick" + 0.032*"size" + 0.030*"door" + 0.023*"bit"
46	6,900	0.125*"tool" + 0.116*"hand" + 0.079*"use" + 0.051*"handl" + 0.046*"work" + 0.036*"die" + 0.031*"cut" + 0.028*"need" + 0.024*"make" + 0.023*"press"
47	5,031	0.120*"print" + 0.100*"model" + 0.044*"printer" + 0.032*"object" + 0.030*"use" + 0.027*"plastic" + 0.027*"design" + 0.026*"materi" + 0.021*"filament" + 0.019*"creat"
48	4,132	0.085*"stand" + 0.059*"carbon" + 0.045*"pro" + 0.037*"use" + 0.035*"fiber" + 0.028*"ultim" + 0.020*"instrument" + 0.019*"like" + 0.018*"work" + 0.017*"make"
49	7,204	0.053*"learn" + 0.049*"game" + 0.034*"school" + 0.030*"student" + 0.022*"educ" + 0.022*"program" + 0.022*"skill" + 0.020*"class" + 0.017*"experi" + 0.017*"scienc"

EC.9. Validation of Our Metric of Number of Product Features

We analyze the relationship between the development level of a product and the number of features in its description. For this analysis, we turn to another crowdfunding platform, Indiegogo. On Indiegogo, in addition to providing a textual description of a product, each creator also specifies the product's development stage (concept, prototype, production, or shipping). As a more advanced stage typically indicates a higher number of features (Indiegogo 2021), we fit the same LDA model as above to a corpus of 10,047 product descriptions on Indiegogo and analyze the relationship between the product's development stage and the number of features. Because *product stage* is a categorical variable (*concept*, *prototype*, *production*, or *shipping* stage), we use a multinomial logistic regression. We regress *product stage* on *features*, *category*, ln(*goal*), *delivery time*, and *funding type*. All variables are as defined in our main models, and *funding type* is either flexible or fixed, depending on whether the creator can keep the money raised or not when the goal is not reached.

As Table EC.9.1 summarizes, the number of features in concept-stage campaigns is significantly smaller ($p < 0.01$) than the number of features in prototype-stage campaigns, and the number of features in prototype-stage campaigns is significantly smaller ($p < 0.01$) than the number of features in production-stage campaigns. These results show that the number of features significantly predicts the product stage. (Although the number of features in production-stage campaigns is not significantly smaller than the number of features in shipping-stage campaigns, as shipping-stage campaigns are not launched on Kickstarter, this result is not important for our analysis.) As products that are more developed tend to have more features (e.g., Althuizen and Chen 2021), this analysis indicates that our LDA model generates a good proxy for the number of features that products have.

Table EC.9.1 Multinomial logistic regression result for Indiegogo.

mlogit (*N*=4153): Factor change in the odds of *Product stage*
Variable: Initial development level (standard deviation=8.672)

	b	z	P>z	exp(b)	exp(b*SD)
Prototype vs. *Concept*	0.082	15.998	0.000	1.086	2.038
Production vs. *Concept*	0.108	16.254	0.000	1.114	2.553
Production vs. *Prototype*	0.026	4.572	0.000	1.026	1.252
Production vs. *Shipping*	0.006	0.716	0.474	1.006	1.057
Shipping vs. *Concept*	0.102	11.778	0.000	1.107	2.415
Shipping vs. *Prototype*	0.020	2.452	0.014	1.020	1.185

b = raw coefficient z = z-score for test of b=0
exp(b) = factor change in odds for unit increase in initial development level
exp(b*SD) = change in odds for standard deviation increase in initial development level

EC.10. Robustness Checks

In this section, we provide results of probit and IV models for robustness checks that we discuss in §3.4.

Table EC.10.1 Spline regressions for second-stage estimations in IV models.

	Second Stage of IV Model 1a Existence of comment(s)	Second Stage of IV Model 2 Product improvement	Second Stage of IV Model 3 Campaign success
Initial number of features ($\leq$ 30)	0.092***	0.053***	0.073***
	(0.009)	(0.008)	(0.008)
Initial number of features (> 30)	-0.031	-0.115***	-0.056**
	(0.026)	(0.029)	(0.023)
Residuals	-0.074***	-0.036***	-0.046***
	(0.009)	(0.008)	(0.009)
Controls	Yes	Yes	Yes
Constant	-0.767***	-0.929***	2.292***
	(0.074)	(0.076)	(0.083)
Observations	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table EC.10.2 Equal time periods before and after IV.

	First Stage of IV Models Initial number of features	Second Stage of IV Model 1a Existence of comment(s)	Second Stage of IV Model 1b Number of comments	Second Stage of IV Model 2 Product improvement	Second Stage of IV Model 3 Campaign success
Initial number of features		.121***	.004	.086***	.138***
		(.014)	(.247)	(.014)	(.016)
Initial number of features^2		-.002***	-.006	-.002***	-.003***
		(0)	(.005)	(0)	(0)
Before relaxation of rules	2.758***				
	(.169)				
Residuals		-.047***	.13	-.012	-.046***
		(.014)	(.245)	(.014)	(.015)
Residuals×Residuals		.001***	.011	0*	.001**
		(0)	(.015)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-1.838***	-.884***	-16.171***	-1***	2.19***
	(.44)	(.1)	(2.915)	(.105)	(.11)
Wald χ^2	6797.80	2874.15	601.55	1045.94	2693.71
R^2 or Pseudo R^2	0.416	.148	0.071	.061	.189
Observations	11,764	11,764	11,764	11,764	11,764

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.3 Treating canceled campaigns as failed campaigns.

	First Stage of IV Models Initial number of features	Second Stage of IV Model 1a Existence of comment(s)	Second Stage of IV Model 1b Number of comments	Second Stage of IV Model 2 Product improvement	Second Stage of IV Model 3 Campaign success
Initial number of features		.121***	.134***	.756***	.103***
		(.014)	(.008)	(.142)	(.007)
Initial number of features^2		-.002***	-.002***	-.006**	-.002***
		(0)	(0)	(.003)	(0)
Before relaxation of rules	3.562***				
	(.134)				
Residuals		-.065***	-.538***	-.028***	-.037***
		(.007)	(.138)	(.007)	(.008)
Residuals×Residuals		.001**	.001	.001***	0*
		(0)	(.006)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-2.738***	-.879***	-14.421***	-1.139***	2.167***
	(.366)	(.079)	(1.883)	(.075)	(.087)
Wald χ^2	11116.64	4251.10	1094.22	1479.75	4058.39
R^2 or Pseudo R^2	0.399	.145	0.057	.051	.178
Observations	21,184	21,184	21,184	21,184	21,184

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.4 When the number of topics is set to 40 in LDA Model.

	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features		.157***	.557***	.114***	.147***
		(.01)	(.184)	(.009)	(.01)
Initial number of features²		-.003***	.003	-.003***	-.003***
		(0)	(.006)	(0)	(0)
Before relaxation of rules	3.245***				
	(.109)				
Residuals		-.073***	-.524***	-.04***	-.043***
		(.01)	(.143)	(.009)	(.01)
Residuals×Residuals		.001**	.002	0	.001**
		(0)	(.01)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-1.086***	-1.046***	-16.591***	-1.138***	2.078***
	(.334)	(.079)	(1.93)	(.075)	(.086)
Wald χ^2	9343.21	2750.55	2187.54	1067.58	3619.10
R^2 or Pseudo R^2	.411	.156	.073	.045	.207
Observations	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.5 When the number of topics is set to 60 in LDA Model.

	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features		.128***	.632***	.095***	.116***
		(.008)	(.119)	(.008)	(.008)
Initial number of features²		-.002***	-.004	-.002***	-.002***
		(0)	(.003)	(0)	(0)
Before relaxation of rules	3.74***				
	(.139)				
Residuals		-.067***	-.474***	-.028***	-.041***
		(.008)	(.125)	(.007)	(.009)
Residuals×Residuals		.001***	.003	0**	0***
		(0)	(.006)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-3.2***	-.818***	-15.982***	-1.071***	2.266***
	(.385)	(.077)	(1.915)	(.083)	(.083)
Wald χ^2	7867.45	2498.67	1800.05	1682.61	3638.74
R^2 or Pseudo R^2	.391	.153	.073	.062	.203
Observations	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.6 When the threshold is set to 8 while counting the number of topics.

	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features		.125***	.505***	.106***	.116***
		(.008)	(.133)	(.007)	(.008)
Initial number of features²		-.002***	0	-.002***	-.002***
		(0)	(.003)	(0)	(0)
Before relaxation of rules	4.031***				
	(.145)				
Residuals		-.059***	-.429***	-.04***	-.035***
		(.008)	(.116)	(.006)	(.008)
Residuals×Residuals		0	0	0	0*
		(0)	(.01)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-1.928***	-.989***	-16.514***	-1.181***	2.126***
	(.415)	(.078)	(1.944)	(.09)	(.085)
Wald χ^2	9497.34	2711.88	2150.86	1301.31	3583.78
R^2 or Pseudo R^2	.411	.155	.072	.053	.207
Observations	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.7 When the threshold is set to 12 while counting the number of topics.

	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features		.157***	.736***	.123***	.143***
		(.01)	(.149)	(.01)	(.01)
Initial number of features²		-.003***	-.004	-.003***	-.003***
		(0)	(.004)	(0)	(0)
Before relaxation of rules	3.043***				
	(.115)				
Residuals		-.082***	-.575***	-.046***	-.049***
		(.01)	(.151)	(.01)	(.01)
Residuals×Residuals		.001***	.005	.001*	.001***
		(0)	(.009)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-2.364***	-.851***	-16.06***	-1.213***	2.241***
	(.325)	(.077)	(1.913)	(.083)	(.084)
Wald χ^2	8324.12	2464.41	1835.97	1487.08	3503.61
R^2 or Pseudo R^2	.389	.153	.072	.057	.204
Observations	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.8 Control for competition in the first week of each campaign.

	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2
	Existence of comment(s)	Number of comments	Product improvement
Initial number of features	.135***	.487***	.104***
	(.009)	(.134)	(.007)
Initial number of features²	-.002***	-.003	-.002***
	(0)	(.004)	(0)
Residuals	-.065***	-.357***	-.032***
	(.008)	(.122)	(.006)
Residuals×Residuals	.001**	.003	0
	(0)	(.007)	(0)
Competition	-.506*	-14.522***	-.499*
	(.262)	(4.437)	(.259)
Controls	Yes	Yes	Yes
Constant	-.862***	-14.876***	-1.107***
	(.076)	(1.671)	(.082)
Wald χ^2	2496.93	2009.62	1476.94
R^2 or Pseudo R^2	.154	.073	.057
Observations	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*

Table EC.10.9 Excluding Topics 4, 11, 12, 18, 19, 20, 21, 32.

	First Stage of IV Models	Second Stage of IV Model 1a	Second Stage of IV Model 1b	Second Stage of IV Model 2	Second Stage of IV Model 3
	Initial number of features	Existence of comment(s)	Number of comments	Product improvement	Campaign success
Initial number of features		.153***	.674***	.112***	.139***
		(.01)	(.152)	(.008)	(.01)
Initial number of features²		-.003***	-.003	-.003***	-.003***
		(0)	(.005)	(0)	(0)
Before relaxation of rules	3.242***				
	(.115)				
Residuals		-.075***	-.526***	-.033***	-.043***
		(.01)	(.141)	(.007)	(.01)
Residuals×Residuals		.001**	.003	0	.001**
		(0)	(.008)	(0)	(0)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-1.927***	-.928***	-16.397***	-1.108***	2.179***
	(.328)	(.077)	(1.921)	(.09)	(.084)
Wald χ^2	9413.41	2475.20	2097.97	1483.81	3577.26
R^2 or Pseudo R^2	.400	.154	0.073	.058	.205
Observations	18,173	18,173	18,173	18,173	18,173

*Nonparametric bootstrap standard errors (100 replications) in parentheses. *** p<.01, ** p<.05, * p<.1*