“Crowds” of Amateurs & Professional Entrepreneurs in Marketplaces

February 13, 2019

Abstract

Digital platform-based marketplaces often have a wide variety of amateurs working alongside professional enterprises and entrepreneurs. Can a platform owner alter the number and mix of market participants? I develop a theoretical framework to show that amateurs emerge as a distinct type of market participant, subject to different market selection conditions, and differing from professionals in quality, willingness to persist on the platform, and in mix of motivations. I clarify how targeted combinations of tweaks to platform design can lead the “bottom to fall out” of a market to large numbers of amateurs. In data on mobile app developers, I find that shifts in minimum development costs and non-pecuniary motivations are associated with discontinuous changes in numbers and types of developers, precisely as predicted by theory. The resulting flood of low-quality amateurs is in this context is associated with equally significant increases in numbers of high-quality products.

Keywords: Amateurs, industrial organization, labor, digitization, long-tail, platforms and marketplaces, complementors, entry and exit, selection and retention, entrepreneurship, minimum viable products, non-pecuniary motivations.

JEL codes: D4, E26, J4, L1, L8, 03.

* Northeastern University and NBER. (kevin@northeastern.edu). I would like to acknowledge helpful comments from seminar participants at the AoM Annual Meetings, the Case Western Workshop on Digital Innovation, Copenhagen Business School, UC Berkeley, Harvard University, the London School of Economics, IESE, INFORMS, the NBER Summer Institute program in Digitization, the NBER Entrepreneurship Working Group, the National Science Foundation digitization meetings, New York University, University College London, the University of Florida Online Platform Conference, the University of Southern California, the University of Michigan, the SMR Special Conference on Open Innovation at UC Berkeley, Stanford University, Warwick University, and the International Industrial Organization Conference and from Carliss Baldwin, Serguey Braguinsky, Henry Chesbrough, Dan Elfenbein, Shane Greenstein, Jorge Guzman, Brent Hickman, John Horton, Justin Johnson, Riitta Katila, Jenny Kuan, Nicola Lacetera, Karim Lakhani, Gwen Lee, Hong Luo, Keld Laursen, Aija Leiponen, Alan MacCormack, Gustavo Manso, Matt Marx, Dylan Minor, Abhishek Nagaraj, Ramana Nanda, Imke Reimers, Tim Simcoe, Chris Stanton, Steve Tadelis, Thales Teixeira, Neil Thompson, Chris Tucci, Eric von Hippel, Joel Waldfogel, Joel West, Youngjin Yoo, and Laurina Zhang. This also benefitted from long-time collaboration on related topics with Lars Bo Jeppesen (CBS) and Milan Miric (USC). Priori Data provided data and industry insights. Research assistance was provided by Richard Paulsen and Gourishankar Seal. Executives at Priori Data contributed proprietary data to this study. I acknowledge the receipt of research funding from Northeastern University and from Harvard Business School in completing this study. All errors remain my own.

1 Introduction

A growing share of trade, interactions, distribution, development, and innovation in the economy now takes place on online platform-based marketplaces (Evans and Gawer, 2016; Parker et al., 2016; Sundararajan, 2016). While a large and growing body of research[1] studies the importance and implications of attracting large *numbers* of suppliers ("developers" or "complementors") onto platforms, much less attention has been devoted to understanding the heterogeneity of these actors. This paper theoretically and empirically studies how the design of platform-based marketplaces shapes the number and types of suppliers working on them.

To appreciate the extraordinary heterogeneity of actors now working on platforms, consider the iconic example of the Apple App Store. Shortly after the launch of the Apple App Store, Der Spiegel observed that "a small industry has developed around Apple's app business, that includes a lot of amateur developers but also incorporates software companies like Zynga and Pangea" (Muller, 2009). Fast Company magazine heralded "...a revival of the hobbyist programmer. Not since the days of the Commodore 64 and Atari 2600 has indie software been sold by such tiny teams of programmers to such massive numbers of consumers" (Stevens, 2011). Now, a "blurry line between professionals and amateurs" (Chip, 2012) exists, with hundreds of thousands of developers ranging from indy developers, entrepreneurial firms, professional enterprises and large corporations to individual hobbyists, learners, hackers, tinkerers, and user-innovators (Chip, 2012; Desai, 2015). Despite incurring $100 in annual platform access fees and on-going capital expenses and opportunity costs, the bulk of developers persist in developing and offering products without any reasonable expectation of earning revenues—much less a positive net income. (The mode of product revenues on the App Store is zero.)

This mix of amateurs and professionals is hardly uncommon and can be seen, for example, in cases of Youtube, Shutterstock, Etsy, Kindle Self-Publishing, and popular podcasting platforms. At the same time, there are cases of platforms dominated by unpaid amateurs, as in cases of add-ons for Mozilla's Firefox web browser, mods of early PC game engines, or in early histories of Youtube or Soundcloud. In other instances—such as Amazon, Alibaba or Uber—suppliers on platforms are predominantly professionals, at least in the sense of expecting to earn some measure of income from their work. Differences in the mix can also be observed in the very same industry, such as corporate versus amateur user-generated video sharing platforms (Ching, 2016).

In this paper, I theorize the conditions determining the number and types of suppliers working on a marketplace, and how to alter this number and mix by specific combinations of changes to platform design. I do so by analyzing a general theoretical framework of selection and retention, consistent with essential features of standard theories in separate streams of work on market entry,

[1] E.g., Rohlfs (1974), Katz and Shapiro (1985), Farrell and Saloner (1986), Arthur (1989), Cusumano (1992), Langlois (1992), Church and Gandal (1992), Bresnahan and Greenstein (1999, 2014), Gawer and Cusumano (2002), Schilling (2002), Rochet and Tirole (2003), Shankar and Bayus (2003), Suarez (2004), Parker and Van Alstyne (2005), Hossain et al. (2011), Zhu and Iansiti (2012), Cennamo and Santalo (2013), Seamans and Zhu (2013).

entrepreneurial selection, and platform adoption (e.g., Bresnahan and Reiss, 1991; Hamilton, 2000; Manso, 2016; Church and Gandal, 1992; Rochet and Tirole, 2003; Parker and Van Alstyne, 2005). Predictions of the theory will also be shown to be precisely consistent with empirical patterns on the Apple App Store.

The theoretical analysis begins by presuming there is some distribution of potential entrants who differ in "quality" (ex-ante expected ability to generate income) and non-pecuniary payoffs. Thus, the broad spectrum of potential entrants—i.e., hackers, students, hobbyists, funded start-ups, tinkers, learners, user innovators, etc.—can be summarized in just two dimensions. Non-pecuniary payoffs, strictly speaking, may include any deviations from standard risk-neutral preferences; but, the motivation and empirics here will focus on factors such as intrinsic motivations, learning, enhancing reputation or career, motivations to use one's own innovations, to work with autonomy, etc..

In solving for conditions determining selection onto the marketplace, I find there is not just one set of selection conditions, but there are two. The first solution (selection condition) delivered by the framework follows the usual idea that a wide range of professional enterprises and entrepreneurs will join so long as they exceed a minimum quality threshold. Most dimensions of platform design will shape this minimum quality threshold, through effects on income, costs, or non-pecuniary payoffs. (Professionals can only be made not to join a platform where the prospect of profit is removed, as when a platform owner proscribes charging its users, where intellectual property is foregone when participating on the platform, where there are merely few users, and so forth.)

The more novel finding is a second solution or selection condition that derives from the same analytical framework describing market participation. This second selection condition relates to a narrower combination of platform design choices. Where the bare minimum cost required to develop a viable product falls below non-pecuniary payoffs (for at least some suppliers), the minimum quality threshold described above no longer applies. It is in this sense that the "bottom-falls-out" of the market to amateurs.

Six predictions follow (with the market, as a whole, as the unit of analysis): (1) dropping minimum development costs in the presence of non-pecuniary payoffs will, at some point, produce a discontinuous response in numbers and types of developers, as the bottom-falls-out and the second selection condition kicks-in; (2) it is only the minimum of costs that is salient here, not other aspects of cost structure; and (3) the level of minimum costs at which the bottom-falls-out will shift upwards or "to-the-right" with a general boost in non-pecuniary payoffs. The developers joining as the bottom-falls-out ("amateurs") will be distinguished by their (4) quality, not adhering to any minimum threshold; (5) persistence, continuing despite low quality and lack of commercial success; and (6) mix of motivations, dominated by non-pecuniary payoffs, while expecting to derive little or negative income. (NB. An amateur with little ex-ante expectation of income might still turn out to be successful ex-post—perhaps even sufficiently so to transform to a professional.)

The ideal experiment to test these six predictions would compare the number and types of developers who choose to participate on identical marketplaces facing identical conditions, while experimentally manipulating minimum development costs and general levels of non-pecuniary motivations. To record any possible discontinuities or nonlinearities, minimum development costs need to be varied over a wide domain. It is sufficient to observe general differences in non-pecuniary motivations as just two levels. Meaningful measures of key theoretical concepts need to be captured over a large population of suppliers on platforms. Such an ideal experiment is not practicable in typical platform industries, as such industries will tend to consolidate around a few differentiated platforms, with little opportunity for precise and uniform measurement. Alternative research designs, such as exploiting one-off platform policy or design changes over time do not provide sufficient variation in minimum development costs.

A research design coming sufficiently close to the ideal, used here, exploits within-platform differences across precisely-defined submarkets. Studying within-platform variation serves to hold constant many of the determinants of marketplace participation, including general platform characteristics and external market conditions. (For examples of analogous research designs that exploit exogenous differences across markets to test theory see, for example, Bresnahan and Reiss (1991), Sutton (1991, 1998), Berry and Waldfogel (1999, 2010), and others.)

To carry out the analysis, I assemble a novel combination of three data sets to study the population of app developers on the US Apple App Store. I exploit within-platform variation across 503 subcategories or submarkets of products (precisely defined by machine learning algorithms and human analysts). The 503 submarkets include, for example, 3,759 weather apps, 528 hockey games, 273 heart monitors, 543 hypnosis apps, 1,643 games on fishing, and 1,543 Sudoku games, and so on.

At the highest level, testing the six predictions essentially involves comparing numbers and types of developers across subcategories, while determining whether these differences are systematically associated with variation in minimum development costs and non-pecuniary motivations—in ways precisely predicted by the theory. In assessing consistency with predictions, it is convenient that tests are not merely a matter of "signs and significance." Instead, the predictions are many (i.e., six); the predictions are precise (e.g., nonlinearities, shifts, etc.); the predictions differ from regular textbook predictions of market outcomes; and—most important—seemingly small differences in market conditions are predicted to be associated with radical differences in market outcomes. Therefore, the tests for consistency with the theory here should hardly be subtle or ambiguous.

Also convenient here, it is possible to draw econometric comparisons across closely similar product submarkets. The 503 subcategories fall into 43 broader categories. Thus, within the category of "Arcade Games," it is possible to directly compare outcomes across similar subcategories such as "Arcade Ball Games" and "Classic Arcade Games," to determine whether differences in minimum development costs relate to stark differences in market participation. For certain tests, it is also appropriate to draw comparisons, not just across submarkets, but also across developers holding the very same rank in their respective subcategories. This allows the analysis the analysis to directly measures whether developers at rank 1, 2, 3, and so on in, say, Arcade Ball Games behave differently from developers at those ranks in Classic Arcade Games—and whether these differences are precisely consistent with theory.

A vital element of the empirical research design is to devise meaningful measures of theoretical concepts across a population of 192,372 developers, 693,541 apps, 503 subcategories, and 43 categories. At the heart of the research design is a novel strategy for capturing differences in minimum costs (complexity, development challenge) across product subcategories. I argue and then I demonstrate that differences in minimum development costs across product types can be proxied by differences in the size of very smallest apps across product types. Or, rather, at least that is true in the Apple App Store, given its rather particular institutional characteristics, and where it is possible to measure product subcategories with great precision.

Patterns in the data each found to be precisely consistent with the six predictions. An interplay of minimum development costs and non-pecuniary motivations is associated with more than doubling the number of developers in each subcategory. The baseline average number of developers per subcategory is 362; where the discontinuity is crossed, there are on average 856 developers. The developers added with the discontinuity are not just low or incrementally lower quality; the vast majority of added developers are of most inferior observable quality. Despite this lowest quality, they persist longer on the platform. Consistent with the relative importance of non-pecuniary motivations, these developers engage in higher levels of development activity than might otherwise be expected (measured as product versioning). Survey evidence corroborate the relative importance of non-pecuniary payoffs. (In a supplemental analysis of the available data, I also find that the bottom-falling-out to great numbers of amateurs is is associated with roughly double the number of highest-quality products.)

The theoretical findings and empirical corroboration in this paper contribute to growing research on platform design decisions, including those involving complementors' cost structure (e.g., provision of development tools, documentation, high powered APIs, platform access fees), along with platform design choices shaping platform complementors' non-pecuniary payoffs (e.g., public profiles of complementor accomplishments, social interaction forums, etc.), and prospects of earning income (e.g., licensing terms, protection of intellectual property rights, etc.). See, for example, Ceccagnoli et al. (2012), Claussen et al. (2013), and Ghose et al. (2014). Here, I identify a specific combination of tweaks that grant access to amateurs.

The paper builds on well-established findings that online developers often work for reasons other than just seeking income (e.g., Lakhani and Von Hippel, 2003; Lakhani and Wolf, 2005; Jeppesen and Frederiksen, 2006; Lerner, et al. 2006 ; Roberts, et al. 2006; Wu et al., 2007; Howison and Herbsleb, 2011; Von Krogh et al., 2012; Agrawal, 2014; Lei, et al., 2016; Xu, 2016; von Hippel 2017.) The analysis here shows that the overall "bundle" of non-pecuniary motivations can interact with platform design in ways that lead to radical changes in market participation.

Research outside of platforms and digital contexts has previously considered links between supplier heterogeneity, non-pecuniary motivations, and product and labor market outcomes (e.g., Caves, 2000; Hamilton, 2000; Scott Morton and Podolny, 2002; Stern, 2004; Roach and Sauermann, 2010; Sauermann and Cohen, 2010; Åstebro et al., 2014). This paper goes further, however, in clarifying differences between amateurs (hobbyists, learners, and fanatics) willing to participate and incur personal expenses without any reasonable prospect or expectation of positive income versus those workers and entrepreneurs who might simply be willing to accept lower-than-regular returns. The framework used here builds most closely on the appealing general framework of Manso (2016), where I work out implications of heterogeneity and endogenous costs to reveal that amateurs can co-exist with professionals.

Closely-related research in the economics of information systems and digitization has previously made broad observations of historical time-trending towards digitization and lower with implications for growing numbers of suppliers (e.g., Brynjolfsson, et al. 2011; Waldfogel, 2018; Waldfogel and Reimers, 2015). The analysis here attempts to more directly theorize and measure the role of costs, showing that "low costs" on their own are not sufficient to grant access to large numbers of amateurs, and it is a specific component of cost structure that is most relevant.

More generally, the findings provide an explanation for why extraordinarily large numbers of developers choose to participate on platform-based marketplaces in competitive equilibrium, while persistently making negative income and having no reasonable expectation of doing better. This explanation does not rely on presuming that tens or hundreds of thousands of suppliers are systematically irrational, or profoundly risk-seeking—nor that network effects are so potent as to justify hundreds or thousands of developers making roughly the same product. Further, while I find that modern online platform-based marketplaces can be analyzed within a standard framework, this is not to say that these markets function the same way as traditional goods markets of old.

The paper proceeds as follows. Section 2 develops the analytical framework leading to six predictions. Section 3 describes key elements of the empirical research design and the data. Section 4 presents the main empirical results. Section 5 provides the supplemental analysis on products. Section 6 summarizes, concludes, and discusses contributions to the literature.

2 Hypotheses Development: Heterogeneous Developers in an Entrepreneurial Marketplace

This section develops predictions concerning the conditions under which amateurs and professionals will choose to participate in a marketplace. This set-up most closely follows that of Manso (2016), while allowing for (i) heterogeneity of agents, and (ii) partially endogenously expenditures in product development and—most important—solving for (iii) two distinct conditions for joining a marketplace.[2]

Time proceeds in discrete periods. In each period, a new unit mass of potential developers arrives (e.g., completing academic training, leaving other employment or leisure, etc.). Each developer lives two periods or phases, $t \epsilon \{1, 2\}$. Therefore, market participation at any time is the sum of phase 1 and phase 2 developers.

In the first "selection" phase, developers choose whether to join the platform. Product development is uncertain and fraction $p \epsilon [0, 1)$ will be revealed to be successful in the sense of earning non-zero revenues (net of any variable costs), $R_{i,1}$ and $R_{i,2}$. Developers also enjoy non-pecuniary payoffs, $\beta_i \epsilon \mathbb{R}^+$, when participating on the platform. (Note, β_i will be treated as a constant in this simple organizing framework; the more general underlying presumption is just that these non-pecuniary payoffs are far less responsive to endogenous market outcomes than are income payoffs.) Subscripts i and t index developers and phase.

A developer will have incentives to enter and experiment so long as the following *selection condition* of the first phase is met:[3]

$$p\left(R_{i,1} + R_{i,2} + 2\beta_i\right) + (1-p)\left(\beta_i + W_{i,2}\right) \geq W_{i,1} + W_{i,2}. \tag{1}$$

The above expression presumes that exit costs are sufficiently low that developers return to their outside option if they are unsuccessful. Note, income R and p and costs W should each be interpreted as values determined in competitive equilibrium. However, developers are small relative to the entire market and R- and p-"takers." Predictions here will, therefore, not depend on strategic interactions or anticipating positive or negative externalities with competitors.

A developer who joined in the first phase will have incentives to continue so long as the following *retention condition* of the second phase is met:

$$R_{i,2} + \beta_i \geq W_{i,2}. \tag{2}$$

Crucially, the vast heterogeneity of developers who might choose to select onto the platform is summarized as differences in quality, $\rho_i \epsilon \mathbb{R}^+$, and differences in non-pecuniary motivations, $\beta_i \epsilon \mathbb{R}^+$. Quality ρ_i together with the chosen level of expenditures and opportunity costs devoted to product development $W_{i,t} \epsilon \mathbb{R}^+$ determine revenues, $R'(W_{i,t}, \rho_i) \geq 0$.

Development costs are partially endogenous, the sum of minimum costs required create a bare minimum viable product, w_{min}, and any additional discretionary quality-improving expenditures, $w_{i,t}$, i.e., $W_{i,t} = w_{min} + w_{i,t}$.[4]

[2] I also note below how the framework allows for externalities among developers, including network effects as are emphasized in the literature on platforms. This point is not essential to deriving results.

[3] NB. The expression simplifies to $pR_i > \frac{(1+p)}{2}W$, equivalent to expression 2 of Manso's (2016) analysis, where R's and W's are constant across phases and non-pecuniary payoffs are zero.

[4] Opportunity costs will also differ across developers and can be understood as reflected in differences in ρ; for given costs, developers with high opportunity costs have less output.

This general set-up will allow the effects of a great number of potential platform design and policy changes to be considered, as below.

<TABLE 1>

Lowering-the-Bar to "Fringe" Professionals. The selection condition (1) defines the part of the ρ and β distribution of developers who will have incentives to join the platform. A first (standard) solution to the selection condition (1) is derived by substituting the partially endogenous cost structure,

$$p\left(R_{i,1}+R_{i,2}\right)-w_{i,1}-pw_{i,2}\geq(1+p)\left(w_{min}-\beta_i\right), \tag{3}$$

and rearranging to show that selection in the first phase requires meeting a *minimum quality threshold*:

$$\rho_i\geq\rho_{min}=\pi^{-1}\left((1+p)\left(w_{min}-\beta_i\right)\right), \tag{4}$$

where $\pi(\rho)\equiv p\left(R_{i,1}^{*}+R_{i,2}^{*}\right)-w_{i,1}^{*}-pw_{i,2}^{*}$, is roughly equivalent to expected income—and is thus increasing in quality, $\pi'(\rho_i)>0$. Developers meeting this minimum quality condition are "professional" in the sense that their willingness to join depends on having some non-zero expectation of revenues. "Marginal" or "fringe" professional developers are those for whom condition 4 is binding.

The retention condition (2) is re-written, below, in a similar order to expression (3), above, to clarify that those joining in the first phase will continue in the second phase, if they were successful:

$$R_{i,2}-w_{i,2}\geq w_{min}-\beta_i. \tag{5}$$

It follows that just fraction p of developers from $t=1$ are retained at $t=2$. The total total number of professionals in the marketplace at any given time is therefore $1+p$ times the number selecting onto the platform in the first phase. (It follows, too, that second-phase developers will have higher incentives to "scale-up" and invest in product quality, as they are by then certain of their returns. By contrast, those in the first phase will realize returns only with probability p, drastically reducing marginal returns to investment.)

These results therefore capture usual intuitions and standard results that participation by professional enterprises and entrepreneurs depends on meeting a minimum quality threshold and attaining commercial success. Changes to each dimension of platform design listed in Table 1 will *incrementally* move this threshold up or down. (Particulars depend on functional form assumptions.)

For example, reducing any part of cost structure (lower w_{min}, greater R'_w,) or taking action to boost non-pecuniary payoffs (β) or boosting the income opportunity (R,p) will generally "lower-the-bar"—adding greater numbers of lower-quality marginal or fringe professionals. Panel I of

Figure 1 shows the minimum quality threshold. The minimum quality threshold in this depiction is first flat and then descends to the right, as this example presumes some fraction of developers has zero non-pecuniary payoffs, while others have some distribution of non-pecuniary payoffs. Panel II of Figure 1 provides an example of lowering the minimum quality threshold or "lowering-the-bar."

The Bottom-Falling-Out to Amateurs. Selection condition (1) does not just have just one solution. This condition is also necessarily satisfied if the following is true:

$$w_{min} \leq \beta_i. \tag{6}$$

Consider that $w_{min} \leq \beta_i$ implies that the left hand side of condition 1 will—at minimum—go to zero. This occurs if quality is sufficiently low that discretionary investments are unproductive and set to zero (i.e., an instance with zero revenues and zero endogenous expenditures). But, where $w_{min} \leq \beta_i$, the right hand side of condition 1 is at most zero.

This solution follows the simple intuition that amateurs are willing to join without any necessary expectation of revenues, so long as their non-pecuniary motivations exceed opportunity costs, or $2\beta_i \geq W_{i,1} + W_{i,2}$ (cf. expression 3). Therefore, the "marginal amateur" is the developer willing to join, even when having no incentives to make discretionary investments beyond the minimum, i.e., $W_{i,1} = W_{i,2} = w_{min}$, or $w_{min} \leq \beta_i$. This means that market selection acts through an entirely different mechanism for developers conforming to this second solution. Let us, for the moment, refer to developers conforming to the second solution as "amateurs," a point I will return to and justify, below.

The implications for retention are even more radical than those for selection.Whereas only fraction p of professionals entering and experimenting at in the first phase are retained in the second phase; *all* developers conforming to this second solution will persist in the second phase so long as $w_{min} \leq \beta_i$ remains true. The total number amateurs at any point in time is $2\times$ the number joining in the first phase (not just $1 + p$ this number).

A "Blurry" Continuum with Nonetheless Distinct Groups. Consider a thought experiment, reflected in Figure 1. A platform does not initially create conditions for amateurs to join (e.g., Panels I and II of Figure 1). Only professionals join and do so in relation to a minimum quality threshold. Now imagine the platform owner drops minimum development costs further (e.g., reducing fixed access, providing more technical support, etc) to a point where the second selection condition can be satisfied and the bottom-falls-out, $w_{min} \leq \beta_i$, for developers with higher β's, as in Panel III of Figure 1.

Note too that those meeting the condition $w_{min} \leq \beta_i$ will include some of those who would have have joined even without a fee change, in the top right hand corner of the distribution of types along with the potentially many more developers who join now below the earlier minimum quality threshold (Panel II versus III of Figure 1). Some of those developers therefore will meet both conditions (4) and (6). So are they now amateurs or professionals?

To distinguish professionals from amateurs in this top-right corner of Panel III, we might simply note that those with sufficiently high quality will endogenously choose levels of product development that exceed their non-pecuniary payoffs, $w_{min} + w_{i,t} < \beta_i$, meaning their participation depends on some expectation of income. Thus, while amateurs do not have a quality floor or minimum threshold, they do have a quality ceiling:

$$\rho_{max} < w^{-1}\left(\beta_i - w_{min}\right). \tag{7}$$

<FIGURE 1>

Thus, the seeming vast continuum of types of developers and "blurry line" separating amateurs from professionals and amateurs (see Introduction) follows from the two solutions and selection regimes being immediately touching in type space in ρ and β. This immediately follows from the $\pi^{-1}(-)$ term in the minimum quality threshold for professionals (4) being mathematically undefined for negative arguments, i.e., $w_{min} \leq \beta_i$. But, this is *precisely* the threshold at which the marginal entrepreneur joins. Therefore, the two selection conditions must "touch."

Therefore, on the one hand, it should be difficult to precisely identify amateurs from professionals if presented with a line-up of fringe professionals and amateurs. Nonetheless, they remain distinct groups in relation to the selection condition to which they conform.

The Importance of (Low) Minimum Costs vs. (High) Non-Pecuniary Motivations Despite the seeming symmetry of (low) minimum costs and (high) non-pecuniary motivations in expression (6), these factors may differ in importance. For example, as captured by the subscript i on the non-pecuniary motivations term β, policies intended to shape non-pecuniary payoffs might more likely generate subjective effects than will policies and design choices affecting cost structure.

As an empirical matter, digital platforms and associated digital development tools are widely appreciated to have reduce costs of market access and development by orders of magnitude to historically low levels (e.g., Waldfogel, 2018). By contrast, non-pecuniary motivations are hardly unique or exclusive to just online digital marketplaces (e.g., Hamilton, 2000; Scott Morton and Podolny, 2002; Stern, 2004; Åstebro et al., 2014). Further, one might imagine it to be far more difficult to increase their value by orders of magnitude, analogous to the fall in costs. The statement of predictions will reflect this asymmetry.

2.1 Summary of Predictions

Key predictions following the above arguments are illustrated in Figure 2 and summarized below:

(1) Reductions in bare minimum development costs required to make a product will, at some point, produce discontinuous changes in numbers and types of developers as the bottom-falls-out (i.e., the second selection condition kicks-in).

(2) Variation in minimum development cost is salient to these effects, not other aspects of cost structure.

(3) A general boost in levels of non-pecuniary payoffs will shift upward ("to the right" in Figure 2) the level of minimum costs at which the bottom-falls-out.

<FIGURE 2>

The latter three predictions relate to the particulars types of developers who join as the bottom-falls-out:

(4) *Quality*: Whereas professionals must have quality above some minimum threshold, quality of amateurs is unbounded from below (where quality relates to ex-ante expectations of income, in excess of opportunity costs).

(5) *Persistence*: Whereas professionals' participation is conditional on quality and on encountering commercial success, amateurs persist despite their low quality and lack of success (inasmuch as condition (6) holds).

(6) *Motivations*: Amateur developers must have non-pecuniary payoffs and these will necessarily dominate income payoffs (if only because income is expected to be little or negative).

The remainder of the paper is devoted to empirically testing these predictions.

3 Empirical Research Design

The empirical analysis tests whether the six predictions of Section 2.1 are precisely consistent with patterns found in a representative platform-based marketplace, the Apple App Store.[5] (A supplemental section also reports implications of the bottom-falling-out on the products that become available.)

The ideal experiment to test these six predictions would compare the number and types of developers who choose to participate on identical marketplaces facing identical conditions, while experimentally manipulating minimum development costs and general levels of non-pecuniary motivations. Minimum development costs need to be manipulated over a wide domain (to capture

[5] A large and growing set of empirical studies have studied mobile app developers. An incomplete list, for example, includes: Gans (2012), Ghazawneh and Henfridsson (2013), Garg and Telang (2014), Ghose and Han (2014), Lee and Raghu (2014), Bresnahan et al. (2014, 2015), and Claussen, et al. (2013), Yin et al. (2014).

nonlinearities or discontinuities); non-pecuniary payoffs can be independently varied at just two levels. The research also needs to observe measures of numbers and types of developers.

A research design coming closest to this ideal, and used here, exploits within-platform differences across precisely-defined submarkets, exploiting differences in minimum development costs and non-pecuniary payoffs across submarkets. The research design here exploits variation across 503 market niches or product subcategories. The key challenges of research design are (i) to make controlled comparisons across comparable markets; (ii) to adhere to high demands of measurement; and (iii) to econometrically isolate and exploit meaningful variation in minimum costs and non-pecuniary motivations. These points are discussed in this section.

3.1 Data Set

The data used here come from three sources. A first source is machine-collected data on the population of 192,372 developers and 693,541 apps in mid-2013 on the US Apple AppStore. The data include all readily-observable facts to App Store users, including app titles and developer names, prices, sales ranks, version numbers, file sizes, and user ratings. Focusing on data from 2013 has the advantage of allowing app-level revenues to be estimated from rankings and pricing data, using a procedure described by Garg and Telang (2012).[6] This period is also one of stable platform growth, five years after the App Store's launch.

These data were matched with a second data source provided by market analyst firm, Priori Data, that categorize each app by machine-learning and expert judgment by the Priori Data'sanalyst team. The apps divide into 43 categories, which break down into a further precise 503 subcategories. For example, within the 693,541 apps, there are 155 aliens and space invader games, 2,246 arcade ball games, 2,517 board games, 1,830 brain teasers, 1,002 advanced calculators, 1,287 alarm clocks, 921 astrology-related apps, and 600 apps related to baby names. Main variables are defined in Table 2.

<TABLE 2>

A third data source is a unique survey data set that is described further within the analysis. These data provide detailed information on developer motivations. Measures capturing our most important variables, capturing broad subcategory-level shifts in w_{min} and β, are described further, in following subsections.

3.2 Measuring Differences in Minimum Development Costs, w_{min}

The measurement strategy must measure precise concepts from the theory (Section 2), and do so across hundreds of thousands of developers and hundreds of submarkets, beginning with minimum

[6] Apple has since changed its ranking algorithm.

development costs. Here, I exploit differences in the minimum file size (i.e., smallest product) across precisely-defined product types to capture differences in minimum development costs (complexity and development challenge) across the product types.

Of course, generally speaking, file size is hardly ever a useful proxy for development costs. An extensive literature studies determinants of file size and software development costs, linking these things to elegance and parsimony of design, architectural approach, use of external code libraries, compatibility choices, and many other factors (e.g., Verner and Tate, 1992; Pendharkar, 2004).

Nonetheless, the bare minimum size of software can serve as an indication of costs, complexity, and effort in solving a particular software design problem (Albrecht and Gaffney, 1983; Boehm et al. 2000). The following discussion provides both a priori rationalization for why this measurement strategy should be effective (given particulars of the institutional environment) and a posteriori validation that the measurement strategy does in fact work, in a direct inspection of the data. The econometric analysis to follow provides further diagnostic tests and validation. [7]

a. Precise Minimum Requirements—and Minimal Non-Essential Variation in Code. On the App Store, bare minimum products are meaningful products that must meet a precisely defined standard of quality and function. For an app to be admitted to the platform, it must pass a rigorous certification assessment, including machine-based and human inspection. The process ensures an app functions properly and does what it says it will do. For example, the app must not jeopardize the security and proper operation of the system—including long load times, improper use of file systems and storage, or system crashing. The certification process also forbids unnecessary "calls" on the platform that do not reflect a product's essential function. For example, applications accessing a user's GPS location must fully document the explicit purpose relative to the program's function (rather than needlessly harvesting data).[8]

The App Store also limits "inessential" or subjective variation in software design. For this reason, there will likely be little "noise" in file size that does not reflect essential functions. For example, the code must not include open source code or other content subject to external legal obligations. It may not draw upon external code libraries, misused trademarks or logos, or inappropriate content. The code will also reflect essential logic and functions, rather than subjective aesthetic or user interface differences. In particular, all software must adhere to standard "Human Interface Guidelines" (http://sdtimes.com/app-store-review-guidelines-smbs/). Adherence is facilitated by Apple's providing standardized code and frameworks for interface elements such as "scrollers" and menus, rules and templates for using a device's buttons, and other elements encour-

[7] Note: This measurement strategy does not imply (i) costs are a proxy for file size, generally; nor does it imply that (ii) the very smallest product is necessarily the product the in fact required very lowest costs; nor does it imply that (iii) a category with smaller smallest products necessarily has lower costs (this need only statistically be true, not deterministically); nor does this imply (iv) this measurement strategy should be relevant in other contexts or other data sets—particularly data sets without the sorts of precise measures of product types or without a precisely defined minimum set of requirements for minimum products.)

[8] The requirements for approval have been made more stringent over time.

aging standard user interface motifs and principles. It is also the case that smallest products do not contain large stores of content, databases, or sophisticated graphics beyond the bare minimum in fulfilling their essential functions.

b. Differences across Product Types. Direct inspection of differences in *MinFileSize* across subcategories immediately reveals differences related to the complexity and development challenges inherent to different subcategories. Consider, for example, the smallest product within the "Battery Monitors" subcategory, "Battery Charge Counter," offers the simplest kind of function in this subcategory. The app essentially counts instances of internal system operations. To perform this task, the app requires just 0.034 megabytes (MB) of code.

Compare this with the smallest product in "Phone Trackers," called "Phone Locator Locate Anyone," which involves gaining permissions from other users to track GPS locations. While this is a non-trivial function–involving networking and satellite geo-location, the app needs simply call on these functions already built into the platform. Thus, it only requires more code than did the earlier counter, with 0.060 MB.

Smallest products of considerably larger sizes and complexity can be found in the "Hobby" category, such as those in the "Instrument and Singing" subcategory. The smallest product in this category, "RISCy," involves a rudimentary drag-and-drop interface with a limited command-set for users to create a musical composition. The greater interactivity and built-in logical grammar to allow flexibility of interaction and production of sounds requires 0.117 MB, about twice as much code as the phone tracker.

Within games subcategories, the "Tic Tac Toe" subcategory is rather simple given the nature of this game. The smallest product is in fact called "tictactoe." The simple interface and simplest possible underlying logic for a game requires just 0.026 MB. This is slightly smaller than the earlier "Battery Charge Counter," which also performed basic calculations and provided a simple user interface—but, in addition, required calling upon system data.

Considerably larger than "tictactoe" are smallest games in the "Sports Games" category. For example, in the "Golf Games" subcategory, "Perfect Swing" is 0.104 MB of code, involving a greater amount of graphics and calculations. Whereas golf games involve one shooter, a subcategory involving multiple interacting shooters is "Hockey Games." The smallest product in that case, "Iceball," is almost 50 percent larger at 0.152 MB. Other games subcategories, such as "Ships and Sea Battles," part of the "Action Games" category, have larger sizes still, as they involve more complex plot, characters, game contingencies, and graphics. The smallest in this subcategory, "Battle Pirates," is 0.364 MB.

Widening the aperture to provide perspective on the wider range product types, Table 3 present the 43 categories or products, covering the 503 precise categories of product types—including those reviewed above. The table shows that there are typically a large number of individual subcategories (with an average of about 12 subcategories in each category) and considerable variation in thesmallest products within each category. The categories are listed in descending order of the mean size of smallest products appearing in subcategories in each category. Even these broad patterns readily reveal that more complex and challenging development problems appear at the top of the list (i.e., the most sophisticated games categories), and at the bottom of the list are simplest apps (e.g., your daily communication, board games, calendars and clocks, news and information clients) or apps simply invoking on-board functions of the platform (e.g., chat and messaging, news and information clients).

<TABLE 3>

c. Differences within Product Types. Direct inspection of the distribution of file sizes *within* individual precisely-defined product types immediately reveals patterns consistent with minimum file sizes reflecting a minimum technical constraint or development challenge rather than, say, some random distribution of minimum file sizes. The distribution of file sizes *within* subcategories is shown in Panel 1 of Figure 3.

Consistent with this minimum approximating a minimum design constraint, the lower left of the distribution is a dense mass of observations at and just above the minimum file size—as might be expected in a binding technical constraint or minimum coding challenge. The density abruptly falls to zero below this densely distributed minimum. This high density of observations at or just above the minimum can be observed whether examining the full range of file sizes (Panel 1), just up to 10 MB (Panel 2), or even an extreme fine-grained magnification of the distribution just above the minimum file size to only 0.25 MB above this point (Panel 3).

Consider, by contrast, that the right extreme of the distribution is a sparse "thin" right tail of just 1,268 products (0.18 % of all titles) exceeding 1000 MB (one gigabyte). This right tail of largest products features discretionary investments, including those in sophisticated game designs (e.g., "Call of Duty: Black Ops," "Assassins Creed") and those with rich and detailed content (e.g., offline Wikipedia references, offline mapping applications covering entire continents, and complete language learning courses). Whereas the maximum extremes are highly sensitive to whether or not a particular observation is included, the minimum constraint is not, given there are (very) large numbers of observations (very) tightly clustered in the minimum extreme.

<FIGURE 3>

Also, if the minimum file size were some arbitrarily determined random "draw" rather than a meaningful indication or signal of sorts of technical challenges required to produce a minimum working product (as the earlier examples suggest), we might expect the left extreme distributions of file size to appear to be similar—randomly draw low values. However, any comparison of the left extreme of file size distributions—by max, by percentile, by generalized Kolmogorov-Smirnov distributional tests—immediately reveals differences across the 503 product types, rather than just similarly (randomly distributed) smallest apps.[9]

3.3 Measuring Differences in Non-Pecuniary Payoffs, β

Data collected by Miric, et al. (2019) on 809 app developers representing 7,973 apps using a large survey of app developers are useful for distinguishing observations (subcategories) with relatively high or low general levels of non-pecuniary motivations. Table 4 compares motivations self-reported by part-time and full-time developers, ordered by the differences between part-time and full-time responses. Ordering the data in this manner shows that part-time developers place greatest (relative) weight to "it's a hobby," "to learn new skills," "for fun," "to increase my job prospects," "to use the app myself," and "to be part of the app developer community." For full-time developers, the relative emphasis is on "to make an income" or "to be an entrepreneur." Thus, the patterns suggest that part-time developers uniformly emphasize non-pecuniary payoffs relative to full-time developers, while full-time developers uniformly emphasize income and commercial goals relative to part-time developers. This might be taken, too, to suggest that amateurs—likely more greatly represented among part-time developers—tend to place relative emphasis on non-pecuniary payoffs.

As large as the survey sample may be, it remains too small to meaningfully match to the full population. We may nonetheless exploit these survey data to draw inferences across subcategories. This is particularly so as the survey sample maps closely to the overall population across a range of observable characteristics (Miric et al., 2019). In particular, survey respondents in the 58 games subcategories were 13% ($s.e. = 0.05$) more likely to report engaging in development for intrinsic motivations related to developing "for fun" than those developing in non-games subcategories. The analysis will exploit this point, presuming that games elicit greater degrees of non-pecuniary motivations, all else being equal. Important to note, it is certainly possible that games and non-games differ in ways other than just levels of non-pecuniary motivations. This point will be discussed in the analysis.

<TABLE 4>

3.4 Empirical Framework

Before proceeding to the detailed analysis, it is useful to clarify the basic econometric approach. Testing the six predictions (Section 2.1) for the most part involves regressing some market outcome

[9] The comparison of all possible distributions would involve n!(n-1)!/2 or 503!502!/2 comparisons, times the number of comparison tests. My findings here reflect tests across an arbitrary number of distributions. Simple presentation of the distributions across product types also immediately reveals clear differences in the location of minima of file sizes. The left extreme of these file size distributions each appears similar to the shape of data presented in Figure 3, with the exception that the location of the minima is different. (Figure 3 adjusts the distributions to plot the minima at the same location in the x-axis.)

(related to numbers or types of developers) on variation in minimum costs w_{min} and/or general variation in levels of non-pecuniary motivations β. (Note here, the earlier theoretical framework examined differences across individual developers, indexed by i. Here, the relevant unit of analysis is the submarket, indexed by s, where we examine shifts in general levels of non-pecuniary motivations.)

Market outcomes should also be affected by a wide range of other demand- or supply-side market characteristics—i.e., factors affecting *R, p*, or *W*—such as the nature of potential entrants, or market demand for a given kind of app. Let these other factors be summarized by Δ. If the influence of Δ is approximated as linearly separable, and markets are indexed by s, the estimating equation is as follows:

$$Outcome_s = g(w_{min,s}, \beta_s) + \Delta_s + \epsilon_s \tag{8}$$

where *g(-)* describes the key relationship of interest with minimum costs and non-pecuniary motivations, and ϵ is a zero-mean error term.

Potential Challenges to Interpretation & Unbiased Estimation. If variation the measures of w_{min} or β were *only* to reflect these inherent differences across app types and nothing else, then simple regressions of outcomes on these variables would reflect unbiased estimates of *g(-)*. Such a condition would be met, for example, if *MinFileSize* were to reflect a "hard" technical limit or constraint and there were little scope for other sources of variation. If, however, at least some component of variation in measures of w_{min} or β were endogenously determined, simple regressions would be more difficult to interpret and likely to generate biased estimates.

a. *Omitted Variable Bias*: A first of three possible challenges to interpretation and sources of biased estimation of *g(-)* considered here relates to omitted variables and spurious correlation, if uncontrolled variation in Δ is correlated with measures of w_{min} or β.

b. *Reverse Causation and Response to Competition*: A second possible challenge relates to possible simultaneity and reverse causation if w_{min} and β not only affect outcome variables, but the reverse might also be true. For example, increased competitive pressure is routinely to be a primary determinant of chosen

c. *Measurement Error and Multiple "Draws" of Cost Functions*: Third, there is another kind of reverse causation that could equally be framed as a *measurement error* if a gap or error δ exists between the observed smallest app and true minimum, i.e., $MinFileSize_s = MinFileSize_s^{true} + \delta_s$. Further, the error would need to be large and to be correlated with *NumDevelopers*. This could be the case if adding greater numbers of developers leads the error in measuring the true minimum to converge towards zero, as greater numbers of file sizes are "sampled" or "drawn" as more actors are added.

Assessment of these three challenges and details of estimation are discussed in the analysis.

4 Main Results

This section documents evidence precisely consistent with the six predictions (Section 2.1). Subsection 4.1 covers Predictions 1 and 2, related to a discontinuity in market participation and finds that incremental decrease in minimum costs are associated with a discontinuity in market participation. Subsection 4.2 covers Prediction 3 and finds this discontinuity shifts with general levels of non-pecuniary payoffs.

Subsection 4.3 presents evidence consistent with each of Predictions 4, 5, and 6, finding that developers joining with the discontinuity are of lowest observable quality, and yet they persist and accumulate on the platform. Other patterns are suggestive too of the relative importance of non-pecuniary payoffs.

4.1 Minimum Development Costs & Discontinuous Changes in Market Participation

This subsection relates to Predictions 1 and 2 of Section 2.1. Figure 4 describes the the relationship between *NumDevelopers* and the measure of minimum costs, *MinFileSize.* Here, I make no presumptions of its shape nor causal interpretation, using a nonparametric estimator with locally-weighted least-squares to estimate *g(-)* of equation (8) in a fully-flexible manner. Weights are provided by a second-order Epanechnikov kernel. Moving from right to left, reductions in the minimum cost measure are initially associated with a relatively incremental increase in numbers of developers. At lower levels of the minimum cost measure, there is a discontinuity and seemingly even a "kink" in the relationship.

<FIGURE 4>

<FIGURE 5>

I estimate the location of the "kink" with an unconstrained piece-wise linear model, specifying two independent linear curves and allowing the breakpoint to be estimated as a model parameter:

$$NumDevelopers_s = \begin{cases} \alpha^{low} + \beta^{low} \cdot MinFileSize_s + \Delta_s^{low} + \varepsilon_s^{low} \; if \; MinFileSize_s \leq \delta \\ \alpha^{high} + \beta_1^{high} \cdot MinFileSize_s + \Delta_s^{high} + \varepsilon_s^{high} \; if \; MinFileSize_s > \delta \end{cases}, \tag{9}$$

where s indexes product subcategories, and where "low" and "high" denote below and above the breakpoint; α, β, and δ are model parameters to be estimated; ε terms are zero-mean error terms;

and Δ summarizes all other possible demand- and supply-side factors shaping *NumDevelopers*. Therefore, without adding control variables, the regression model error is the sum of Δ and ε. Parameters are estimated via maximum likelihood.

Figure 4 also shows this second set of estimates, along with 95% confidence intervals. The point of discontinuity δ is estimated to be $MinFileSize = 0.063$ ($s.e. = 0.023$). Parameter estimates of the independently estimated segments lead the two segments to "touch" at the discontinuity. Coefficient estimates are reported in model (2) of Table 5. Model (1) provides a simple comparison of means on either side of the breakpoint: There are on average 362 developers "above" or "to the right" of the point of discontinuity ($MinFileSize > 0.063$) and 856 (i.e., 494 + 362) "to the left" or "below" this point of discontinuity. Figure 5 shows comparable estimates using a quadratic specification, finding no evidence of statistically discernible curvilinearity on either side of the point of discontinuity.

a. Omitted Variable Bias? If the variation in $MinFileSize$ only reflects inherent differences in minimum costs (development challenge, complexity) across different product types, the above associations will measure *g(-)*, as in Expression (8). However, as was discussed in Section 3.4, if some component of *MinFileSize* varies endogenously, this variable would then likely be correlated with demand- or supply-side determinants of market participation—implying spurious correlation and omitted variable bias in the NumDevelopers-MinFileSize relationship (Section 3.4).

Category Fixed Effects: I investigate this possibility in model (3) by adding stringent controls to soak-up variation in demand- and supply-side determinant of market participation (Δ in expression 8) with fixed effects for the 43 product category fixed effects. Consistent with the stringency of these controls, adding fixed effects increases variance explained by 65% in terms of increase in the R^2 statistic (i.e., from 0.17 in model (2) to 0.28), while leaving main model coefficients statistically unchanged.

Demand-Side Controls: We can go still further in seeking evidence of omitted variable bias by introducing subcategory revenues as a model control variable to capture any lingering demand-side variation in Δ not already accounted for by the category fixed effects. As reported in model (4), this leaves main coefficient estimates statistically unchanged, again. The coefficient on revenues is, itself, insignificant. (It is positive if the model is re-estimated without fixed effects.)

Supply-Side Controls: Going still further, model (5) adds a control variable for variation in supply-side factors, roughly proxying for share of international suppliers and other possible variation in supply-side composition with the share of developer websites with suffixes other than ".com" (i.e., .co.uk, .ca, etc.). Again, this does not statistically change main coefficient estimates.[10] While

[10]NB. Revenues and international composition of suppliers are themselves endogenous variables. Their inclusion in the model should only be viewed as a means of illustrating robustness. As endogenous variables, they would be an ineffective means of testing non-robustness. To illustrate, consider that including a simple count of international developers would be correlated with *MinFileSize* if only because *MinFileSize* is correlated with the total number of developers.

the results indicate that many factors may determine market participation, these factors are not correlated with *MinFileSize*.

<TABLE 5>

b. Reverse Causation and Response to Competition? Even where we find no evidence of a spurious correlation, it remains possible that reverse causation could explain results, i.e., if NumDevelopers at least in part causes MinFileSize, rather than just the other way around (Section 3.4). Varying *NumDevelopers*, for example, could plausibly alter competition and product development strategies of makers of smallest products, including makers of very smallest products (as in *MinFileSize*).

Controlling Independent Variation in Dependent Variable: If independent variation in *NumDevelopers* somehow were to cause variation in *MinFileSize,* applying stringent controls for independent variation in *MinFileSize* (as in models (3), (4), and (5) of Table 5), might be expected to significantly perturb the coefficient estimated on *NumDevelopers.* In those earlier tests (the same as those applied to seek any evidence of omitted variable and spurious correlation), coefficients remained statistically unchanged.

Testing Similarity Across Small Developers: To more explicitly scrutinize whether in *NumDevelopers* is leading developers of small products to somehow respond by producing smaller products, Table 6 compares the relationship between *NumDevelopers* and various other percentiles of *NumDevelopers* to determine whether smaller percentiles of the file size distribution are similarly smaller where there are many developers. For simplicity of comparison, just first-order linear relationships are reported. The relationship between *NumDevelopers* and *MinFileSize* in model (1) is, of course, large, and negative. Relationships with first, fifth and tenth percentiles—i.e., models (2), (3), and (4)—are quite similar to one another, but opposite in sign to the *MinFileSize* model (1). Therefore, contrary to reverse causation of this kind, there is no similar response across developers of small products.

Comparing Minimum File Size vs. Wider Distributional Results: Apart from the contrast between the (very) sizeable negative relationship between market participation and the minimum (model 1) versus opposite positive small relationship with other percentiles (models, 2, 3, 4), market participation is also statistically unrelated to mean (model 5) and standard deviation (model 6) of the entire distribution of file sizes.

The special salience of the minimum of file size is especially striking when considering that variation in the minimum is not only meaningfully indicative of subcategory differences (Section 3.2) but that this variation is relatively tiny (mean = 0.058 MB; std. dev. = 0.051 MB). By contrast variation in mean file size (mean = 25.3 MB; std. dev. = 24.8 MB) is four orders of magnitude larger. It is precisely this tiny variation in minimum costs that is predicted by theory to lead the bottom to fall out.

Apart from the minimum, the entire file size distribution appears to be endogenous. If we compare the re-estimate coefficients for models (1) through (6) without product category fixed effects, the coefficients are highly unstable, either halving or doubling (as reported in a bottom of Table 6). By contrast, the coefficient on the minimum measure in model (1) is unchanged, whether including fixed effects or not.

Note too, these results are each consistent with the measurement strategy (Section 3.2), which asserts that the minimum of file size in this context reflects inherent differences in product types. To the extent file sizes (above the minimum) contain even a rough signal of discretionary costs and other aspects of cost structure, these contrasting results are also consistent with Prediction 2, which suggested that minimum costs should be most salient in producing discontinuous changes in market participation.

c. Measurement Error and Multiple "Draws" of Cost Functions? A third possible source of bias or difficulty in interpreting results, as outlined in Section 3.4, is the possibility that there is a gap or error between MinFileSize and the true minimum, and this error is large, and the error converges downwards towards zero with higher numbers of "draws" or greater *NumDevelopers*.

Controlling for Independent Variation in "Number of Draws": Inasmuch as the dependent variable might cause a reduction in *MinFileSize* (or, rather, an error component of this variable), we might again expect that controlling for independent variation in *NumDevelopers* should perturb the coefficient on *MinFileSize*. However, just as above, we see no evidence of any statistical change in estimates in models (3), (4), and (5) of Table 5.

Comparing File Size Distribution, Stratified by "Number of Draws": If the minimum of file size reflects a binding constraint, we might expect a high density at and or immediately above the minimum. If instead, *MinFileSize* were to have a large random component and affect the shape of the "lower tail" of observed file sizes, we would expect something closer to a "thin" left tail or even sparse observations in the vicinity of minimum extreme observations. As was reported in Figure 3, the lower extreme distribution is not just "thick," it is the *mode* of the *entire* file size distribution.[11]

Consider too that if there were a gap or error between *MinFileSize* and the true minimum, this gap should tend to be larger in product types with fewer developers. Figure 6 presents the distribution of file sizes within product types, stratified by types with relatively few or many developers—fewer or greater than 362 developers, the mean number of developers above (to the right) of the discontinuity ($MinFileSize > 0.063$). Figure 6 presents three panels with a progressively finer-grained magnification of the lower extreme of the distribution. Again, the graphs reveal thick

[11]This point should not be confused with or related to the observation that firm size distributions often having a distribution where the minimum is the mode, as when the minimum number of employees is one. Here we are dealing with a continuous variable in file size. Absent a meaningful app certification process (Section 3.2), file size should have followed a smooth distribution, possibly even coming close to zero. Instead, we have abrupt stopping points and meaningful variation well above zero to meet the strict certification requirements described earlier.

"left bunching" at *MinFileSize* and an abrupt fall in density below that, consistent with a minimum constraint or cut-off, both for cases of relatively few or relatively many developers.

Remarkably, there are, on average, three products within just 0.0125 MB of *MinFileSize*, or within the one single increment in histograms at the point of *MinFileSize* in the very bottom highest-resolution presentation of the left extreme distribution in Figure 6). This proximity in size of very smallest products is a virtually infinitesimal in relation to the gigabytes separating the wider population of products within any one category.[12]

Re-estimating with Redefined Minimum Measure: The preceding comparison showed that submarkets with many or fewer developers still appear to have a dense concentration of file sizes at the minimum. As an additional assessment of the extent to which results are not driven by gaps and errors in some smaller subset of subcategories, I also re-estimate the models of Table 5 with *MinFileSize* based instead on second, or third, or fourth smallest products (rather than very smallest). This test provides another means of attempting to detect whether random variation in the minimum extremes might somehow be influencing results. These alternative measures give almost precisely identical estimates, as the tight clustering of observations around the minimum means the redefined *MinFileSize* is little changed when dropping very smallest observations.

Assessing the Shape of the Relationship: If *MinFileSize* were the extreme "draw" of a random process, this extreme value should grow more extreme (lower) with more draws (*NumDevelopers*). Further, each successive draw has a lower chance of lowering the extreme. Higher numbers of draws should produce a gradual curvilinear relationship. However, as was earlier reported, in the *NumDevelopers-MinFileSiz*e relationship, the only discernible convexity is related to the point of discontinuity. There is no evidence whatsoever of curvilinearity on either side of the discontinuity, much less a gradual general curvilinearity.

<FIGURE 6>

The preceding results are consistent with the measurement strategy (Section 3.2) having been a success, with patterns consistent with both Predictions 1 and 2.

4.2 Non-Pecuniary Motivations & Discontinuous Changes in Market Participation

This subsection relates to Prediction 3 of Section 2.1. As illustrated in Figure 2, Prediction 3 indicates that a general increase in non-pecuniary payoffs should increase the level of minimum cost at which the bottom-falls-out, shifting it "to the right." To test consistency with this prediction, we exploit the earlier reported fact (Section 3.3) that those building games applications were 13%

[12] To put this into further perspective, had product sizes in each subcategory been perfectly randomly spaced we would expect to see just one app at roughly 2.75 MB intervals; this expected distance in the left extreme might be higher still if the distribution were some unimodal random distribution with "thin tails."

($s.e. = 0.05$) are more likely to engage in development "for fun" than those developing for non-games-related subcategories. This pattern follows to the intuition that there are higher intrinsic motivations to develop and also play games than for other software.

The analysis here, therefore, compares the 58 observations related to games subcategories to those of other subcategories. As shown in Figure 7, the non-linear relationship of the 58 games subcategories is indeed to the right of that for the broader population. The link is estimated flexibly with a non-parametric estimator, in an analogous fashion to the earlier non-parametric estimate in the previous subsection. (The fewer 58 observations, in this case, do not allow the precise location of the breakpoint in a piecewise model to be estimated with precision, as in the earlier subsection.)

<FIGURE 7>

The estimated relationships in Figure 7 show that these cases of higher non-pecuniary motivations are to the right, with the point of discontinuity occurring at higher levels of the measure of minimum development costs. Of course, games and non-games will differ in ways other than just their general levels of non-pecuniary motivations and in ways that could influence numbers of developers, and we must, therefore, practice caution in interpreting these patterns, accepting them only as consistent with Prediction 3.

In interpreting the weight of evidence, however, it is nonetheless encouraging that the absolute numbers of developers in games and non-games subcategories are the same either to the left or right of the discontinuity; it is only the location of the discontinuity that is statistically different. Therefore, in broadest brushstrokes, the factors influencing market participation do not appear to be all that different across games and non-games.

Further, any concern we should have here for omitted variable bias is less about unobserved determinants of overall numbers of developers. The precise, relevant question here is whether omitted variables can alter the way that MinFileSize relates to NumDevelopers. But, the earlier results in Section 4.1 suggest the MinFileSize-NumDevelopers relationship is quite robust and little influenced by unobserved factors. Therefore, it appears unlikely that had there been some way to control for differences across games and non-games (apart from differences in non-pecuniary motivations) this would have upset consistency with Prediction 3.[13]

[13]Another possibility in interpreting the shape of these relationships is the possibility that non-pecuniary motivations and payoffs are larger where there are higher costs and development challenges. In particular, if differences in development challenges posed by the tiniest products in each category also reflected more general differences in development challenges across the entire spectrum of products within a subcategory, then the patterns of discontinuity could be understated here, as low minimum costs would be associated with generally lower non-pecuniary motivations. However, here the relationship between numbers of developers and the minimum of file size is utterly unlike the relationship with other measures of file size.

4.3 Developer Types and Discontinuous Changes in Market Participation

This subsection relates to Predictions 4, 5, and 6 of Section 2.1—and differences developer characteristics that should be expected with the discontinuity in market participation as the bottom-falls-out.

Quality. The earlier theory of Section 2.1 predicts that added developers with the discontinuity should not just be of incrementally lower quality; they should instead be unbounded from below in quality. To assess evidence, the distribution of quality measures for those subcategories on either side of the discontinuity (i.e., greater or less than $MinFileSize = 0.063$) are compared.

Panel 1 of Figure 8 presents the distribution of product quality as evaluated by users. This is evaluated on a 5-point quality scale. Panel 2 shows the simple count of number of user ratings. (While ratings might, of course, be either positive or negative, the number of ratings is highly positively correlated with sales for apps for which these data are available.)

Consistent with predictions, if we compare the products *added* under especially low minimum costs (i.e., the difference in distributions), we see the majority of added products are not simply of lower quality, but of lowest possible observable quality. The discontinuity does not just represent a shift in the distribution of types, but rather the vast majority loading onto the most leftward increment in the histograms. (Analogous plots of product size or numbers of versions released, each arguably capturing some element of quality, follow a similar pattern.)

<FIGURE 8>

Therefore, patterns are consistent with Prediction 4: the discontinuity of market participation is indeed associated with a flood of lowest-quality suppliers. (Patterns of product quality are analyzed further in Section 5.)

Persistence. The earlier theory of Section 2.1 also predicts that developers joining with the discontinuity will be subject to different selection conditions and will be more likely to persist (less likely to exit) despite their low quality and lack of commercial success. To compare long-livedness on the platform, Panel 1 of Figure 9 presents the lifespan of products on the platform, from top to bottom-ranked apps in their respective subcategories.

Rank order of an app within a subcategory is approximated by the rank according to numbers of user ratings. (Numbers of userratings is strongly correlated with sales, For those data for which sales data are available). Panel 1 effectively presents estimates of the relationship between persistence and rank of app, stratified by cases of low versus high minimum costs, on either side of the discontinuity using the following model:

$$DaysOnPlatform_{s,rank} = \begin{cases} f^{low}(Rank) + CategoryFE_s + \varepsilon_s^{low}\ if\ MinFileSize_s \leq \delta \\ f^{high}(Rank) + CategoryFE_s + \varepsilon_s^{high}\ if\ MinFileSize_s \leq \delta \end{cases}, \tag{10}$$

where, again, s indexes product subcategories, "low" and "high" denote the parts of the *MinFileSize* domain above and below the breakpoint, δ is the earlier-estimated parameterized breakpoint of 0.063 MB, category fixed effects are included to minimize subcategory-specific variation and to add precision, and ϵ captures random error and any subcategory-specific variation not already captured by fixed effects. The *f(-)* relationships are estimated non-parametrically using locally weighted least squares with weights provided by a second-order Epanechnikov kernel.

As might be expected, highest-ranked most-successful apps—to the left of Panel 1 in Figure 9—have the longest lifespan, with a general tendency for durations to descend when proceeding rightward, to lower-ranked apps. Of course, the curve extends over a broader domain in the estimates below the point of discontinuity, as there are many more developers in those cases. With low minimum development costs (below the point of discontinuity $\delta = 0.063$ MB), products are statistically relatively longer-lived on the market. Similar patterns are found in Panel 2, where the unit of analysis is developers rather than products.

<FIGURE 9>

Therefore, patterns are consistent with Prediction 5: the discontinuity of market participation is associated with lowest-quality suppliers who nonetheless persist on the platform.

Non-Pecuniary Motivations. The earlier theory of Section 2.1 predicts too that developers joining with the discontinuous increase in market participation will have high non-pecuniary motivations *relative* to income-orientated motivations. It is not possible to directly observe motivations, much less for hundreds of thousands of market participants. Nonetheless, there are two sets of facts that are consistent with this prediction, as below.

The first set of evidence consistent with this prediction was the earlier self-reported survey evidence from developers, presented in Table 4. The motivations of "it's a hobby," "to learn new skills," "for fun," "to increase my job prospects," "to use the app myself," and "to be part of the app developer community" were reported disproportionately by part-time developers, and a disproportionate number of surveyed part-time developers appear in subcategories below the point of discontinuity.

The observational data also provide an opportunity to seek evidence of non-pecuniary motivations and "excess" development activity relative to what economic incentives would predict on their own. Standard theory, for example, suggests that lower-ranked, lower-quality developers should have lower economic incentives to invest in product development.

Panel 1 of Figure 10 present patterns related to both numbers of versions released for a given title and the number of titles or product scope per developer. Panel 1 provides some suggestion that the number of versions released of a given product is somewhat higher among those in subcategories with lower minimum development costs (when comparing similar product types in the same broader category, but which differ slightly in bare minimum required development costs.)

Panel 2 in Figure 10 similarly analogously reports patterns related to number or products or product scope per developer in cases above or below the point of discontinuity where $MinFileSize <$ 0.063 MB, conditioning on both rank and category. In this case, no statistical differences are observed.

<FIGURE 10>

Therefore, notwithstanding the difficulty of observing motivations, both self-reported and descriptive observational evidence suggest patterns consistent Prediction 6.

5 Supplemental Analysis: Product Quality

The main thrust of this paper, as above, was to identify precise conditions shaping the number and mix of developers in platform-based marketplaces. This section exploits the available data to provide a brief supplemental analysis of the quality of products generated in association with the bottom-falling-out to a flood of amateur suppliers. Whereas earlier empirical results on market participation essentially derive from standard and general frameworks, the effect of "democratizing" innovation to large numbers of amateur developers with little ex-ante expectation of commercial success is theoretically ambiguous—and should depend on many factors.

Below the point of discontinuity with low minimum costs, as in Figure 4 ($MinFileSize < 0.063$ MB), where the bottom has fallen out of the market, the average number of developers is 856, 136 percent greater than the average 362 developers with higher minimum development costs ($MinFileSize < 0.063$ MB). The number of products increases by 151 percent to 1404 from 558 where minimum costs are lower than the point of discontinuity. The flood of low(est) quality developers, as the bottom-falls-out in this context, is associated with higher numbers of high-quality products in this context.

As reported in model (1) of Table 7, the number of products with at least 4.5 out of 5 ratings increases from 111 (s.e. = 17.2) by 158 (s.e. = 22.1) to 269, as the bottom-falls-out. This is a 242% increase in numbers of highly-rated products. It remains possible that lowest-quality and least-reviewed apps have high variance and many "false positives" when counting seemingly high-quality apps. Table 7, therefore, re-estimates the effects of the bottom-falling-out on numbers of apps with at least 4.5 out of 5, but counting only those with a minimum number of user ratings. As reported in models (2) through (5), whether sampling on >10, >100, >500, or >5000 ratings, the number of highest-quality products roughly doubles once the threshold is crossed where there is a flood of added suppliers. Even estimating the relationship on numbers of top apps with at least 150,000 user ratings (the 99th percentile), I find many more (437% times) highest-quality products.

Therefore, the bottom-falling-out to large numbers of amateurs with little expectation of commercial success is associated here with roughly double the number of top-quality products. The results are consistent with, say, the experimentation of amateurs generating useful knowledge spillovers for professional developers at the top of the market. It is also possible that (ex-ante) low-quality amateurs who join the market and experiment, turn out to be high-quality developers (ex-post).

These results are consistent with studies presenting evidence of specific cases where "democratizing" innovation appears to create certain advantages in innovation and problem solving (e.g., Jeppesen and Frederiksen, 2006; Jeppesen and Lakhani, 2010; Kittur et al. 2013; West and Bogers, 2014; Waldfogel, 2014; Waldfogel and Reimers, 2015; Aguiar and Waldfogel, 2016, 2018; Lyytinen, et al. 2016; Sauermann and Franzoni, 2015; von Hippel, 2017). Here, the analysis isolates differences in relation to a comparison group to discern the association with amateurs.

<TABLE 7>

6 Summary & Conclusions

Today, metaphors such as "crowds" or "long tail" are often used to attempt to describe the extraordinary heterogeneity of suppliers now often observed on modern platform-based marketplaces. This paper showed that a relatively standard analytical framework could be used to more precisely understand the number and types of suppliers who participate in modern marketplaces—and how specific combinations of platform design choices can shape market participation. The most novel contribution of this general characterization related to identifying conditions under which amateurs would participate in these marketplaces alongside professional suppliers. Six predictions were borne out in data related to the Apple App Store. The research design allowed precise comparisons to be drawn between submarkets for highly similar products while observing variation in specific conditions predicted to be critical determinants of market participation by the theory.

The Bottom-Falling-Out to Amateurs. The analysis here finds that in addition to traditional professional enterprises and entrepreneurs participating in marketplaces, a second and distinct class of market participant—subject to different selection conditions—can also appear under certain circumstances. Whereas traditional market actors select onto the market if they meet a minimum quality threshold, this second group joins in cases in which the minimum costs required to develop a viable product and participate in the market, fall below the value of non-pecuniary payoffs (for at least some suppliers). I refer to this distinct group of market participants identified in the analysis as "amateurs." As I show theoretically, amateurs and professionals may form a continuum in the distribution of suppliers; however, they are discretely distinct concerning whether they are subject to regular market discipline or not.

Within the paper, I enumerate ways in which the relevant platform design influences parameters related to amateur participation. For example, low minimum development costs can be implemented, for example, by design choices such as low access fees and provision especially simple development environment; non-pecuniary payoffs can be influenced by, for example, the nature of the development task, the enablement of public profiles, or social interaction forums.

Under necessary conditions for amateurs, the "bottom-falls-out" in the sense that the minimum quality threshold disappears, and small combinations of the relevant tweaks to platform design can lead to discontinuous changes in numbers and types of suppliers. The empirical analysis confirmed the sharp predictive power of the analytical framework. Patterns in the data were each found to precisely conform to each of the six predictions. On the App Store, the number of developers more than doubled in association with small shifts in minimum development costs and in association with general changes in non-pecuniary motivations. Also consistent with theory, developers entering with these discontinuities were not just of lower quality, but most inferior observable quality. These added developers also persisted and accumulated in number on the platform, despite their low quality and lack of commercial success (and on-going expenditures to remain on the platform). The evidence is also consistent with non-pecuniary motivations playing a relatively high role in the choices and behavior of these developers.

Therefore, the principal thrust of this paper illustrates that much of the variation in heterogeneity we observe across platforms—between those with amateurs or professionals or some mix acting as platform complementors—is in large part influenced by platform design. The theory and empirical corroboration here also indicate that the "long tail" we observe on many platforms is perhaps many times "longer" than would be predicted by conventional theories of market entry (by professional enterprises and entrepreneurs).

Note too, the analysis focused on free and non-discriminatory conditions of marketplace entry and platform access. This situation is typical of many of today's platforms, where vast numbers of developers lead to the use of standard form contracts (i.e., "click to accept"), standard development kits, and standard commercial terms. I did not consider cases where platform access could be directly regulated or indirectly discriminatory. The theoretical and empirical analyses focused on a single marketplace in isolation (holding constant variation in external conditions). Topics remaining to be researched more closely include: competition among platforms involving amateurs, and general implications for product and labor markets, where amateurs are prevalent. Consider, for example, the most basic fact that the suppliers or developers in the model are willing to engage in production, not just a suppressed wages (cf. Caves, 2000 Ch. 4; Stern, 2004), but indeed at a loss.[14] One presumes these developers must maintain some arrangement (outside the model) for

[14] I speculate that the framework and explanation here might also offer a complementary explanation of market

keeping bread on the table.

Why Should a Platform Owner Let the Bottom-Fall-Out to Amateurs? The radical changes in market participation brought about by tweaks in platform design, and the bottom-falling-out naturally raises the question of why a platform owner might choose to implement such conditions. In a supplemental analysis of the available data, I found that the bottom-falling-out to a flood of amateurs is associated with an increase in numbers of high-quality products. For example, comparing narrowly-defined product subcategories with or without conditions to lead the bottom to fall out, I find the average number of very highest-rated apps with at least 500 ratings is considerably higher under conditions allowing the bottom to fall out, going from 4.9 to 7.3 top high-quality apps.

This result is analogous to other studies finding that "democratizing" of innovation can be associated with high levels of innovation. The analysis is notable in focusing particularly on quality, for estimating differences relative to an explicit comparison (control) group (i.e., similar submarkets where the bottom has not fallen out), and the precise focus on effects of amateurs rather than broader notions of "democratizing" innovation to large numbers of innovators, more generally.

There may be many reasons for the increase in quality with the bottom-falling-out to amateurs. For example, it is possible that the experimentation of amateurs could generate useful vicarious learning and knowledge spillovers for professional developers. It is also possible that some fraction of developers joining as low-quality amateurs might nonetheless turn out to be successful ex-post, contributing to high quality. Features of amateurs could plausibly allow them to play a useful role as platform "farm team" or "test kitchen." For example, despite their low quality (little ex-ante expectation of income), they can persist, linger, and learn. Amateurs escape normal pressures of market discipline. Selecting on the basis of characteristics other than quality (expected income) may also allow a wider diversity of innovators. The outsized role of non-pecuniary motivations may lead amateurs to pursue more personal and idiosyncratic or high-variance projects, outside of usual projects with highest-certainty commercial ends. By comparison, higher-quality commercially-motivated entrepreneurs whose quality and capabilities are better honed to attaining commercial success could be less able or inclined to engage in exploratory innovation.

While product development and innovation questions were supplemental to the main analysis here, these are questions of first-order importance for future research. Whereas, on the one hand, we are entering a period of history where, on the one hand, "ideas are getting harder to find" (Bloom et al., 2017) in association with a growing "burden of knowledge" for innovators (Jones,

structure in cultural industries and the arts or among citizen scientists and tinkerer communities, where amateurs abound. Artists and citizen scientists might not simply "pay to be artists" or "pay to be scientists" in the sense of accepting below-competitive wages; rather, they quite literally pay for equipment and with opportunity cost of time, without any reasonable expectation of a small probability high payoff or even a long run chance of a below-competitive wage.

2009); on the other hand, innovation on platforms has never been easier or more accessible to masses of contributors.

REFERENCES

Afuah, Allan, and Christopher L. Tucci. "Crowdsourcing as a solution to distant search." *Academy of Management Review* 37.3 (2012): 355-375.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. "Some simple economics of crowdfunding." *Innovation Policy and the Economy* 14.1 (2014): 63-97.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. Prediction Machines: The simple economics of artificial intelligence. Harvard Business Press, 2018.

Aguiar, Luis, and Joel Waldfogel. 2016. "Even the losers get lucky sometimes: New products and the evolution of music quality since Napster." *Information Economics and Policy* 34: 1-15.

Aguiar, Luis, and Joel Waldfogel. "Quality Predictability and the Welfare Benefits from New Products: Evidence from the Digitization of Recorded Music." *Journal of Political Economy* 126.2 (2018).

Albrecht, Allan J., and John E. Gaffney. 1983. "Software function, source lines of codes, and development effort prediction: a software science validation." *IEEE Trans Software Eng.* SE-9, pp. 639-648.

Arthur, W. Brian. "Competing technologies, increasing returns, and lock-in by historical events." The economic journal 99.394 (1989): 116-131.

Åstebro, Thomas, Holger Herz, Ramana Nanda, and Roberto A. Weber. 2014. "Seeking the roots of entrepreneurship: insights from behavioral economics." *Journal of Economic Perspectives* 28, no. 3: 49-69.

Bailey, Joseph, Gordon Gao, Wolfgang Jank, Mingfeng Lin, Henry Lucas, and Siva Viswanathan. "The long tail is longer than you think: The surprisingly large extent of online sales by small volume sellers." (2008).

Baldwin, Carliss, and Eric Von Hippel. "Modeling a paradigm shift: From producer innovation to user and open collaborative innovation." *Organization Science* 22, no. 6 (2011): 1399-1417.

Belenzon, S. and Schankerman, M., 2015. “Motivation and Sorting of Human Capital in Open Innovation.” *Strategic Management Journal*, 36(6), pp.795-820.

Benkler, Yochai. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006.

Berry, Steven, and Joel Waldfogel. "Product Quality and Market Size." The Journal of Industrial Economics 58, no. 1 (2010): 1-31.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. Are ideas getting harder to find?. No. w23782. National Bureau of Economic Research, 2017.

Boudreau, Kevin, and Lars Jeppesen. "Unpaid crowd complementors: The platform network effect mirage." *Strategic Management Journal* 36.12 (2015): 1761-1777.

Bresnahan, T. F., & Greenstein, S. (1999). Technological competition and the structure of the computer industry. The Journal of Industrial Economics, 47(1), 1-40.

Bresnahan, Timothy, and Shane Greenstein. 2014. "Mobile computing: The next platform rivalry." *American Economic Review* 104.5: 475-480.

Bresnahan, Timothy F., Jason P. Davis, and Pai-Ling Yin. "Economic value creation in mobile applications." *The changing frontier: Rethinking science and innovation policy*. University of Chicago Press, 2014. 233-286.

Bresnahan, Timothy, Joe Orsini, and Pai-Ling Yin. Demand heterogeneity, inframarginal multihoming, and platform market stability: Mobile apps. Mobile Innovation Group Working Paper, 2015.

Bresnahan, Timothy and Peter Reiss. 1991. “Entry and Competition in Concentrated Markets,” *Journal of Political Economy*, pp. 977- 1009.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester. 2011."Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales." *Management Science* 57.8: 1373-1386.

Brynjolfsson, Erik, Yu Hu, and Michael D. Smith. 2003. "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers." *Management Science* 49.11: 1580-1596.

Boehm, Barry, Chris Abts, and Sunita Chulani. 2000. "Software development cost estimation approaches—A survey." *Annals of software engineering* 10.1-4: 177-205.

Caves, R. E. (2000). Creative industries: Contracts between art and commerce (No. 20). Harvard University Press. Chicago

Ceccagnoli, M., Forman, C., Huang, P. and Wu, D.J., 2012. Cocreation of Value in a Platform Ecosystem! The Case of Enterprise Software. *MIS Quarterly*, pp.263-290.

Cennamo, Carmelo, and Juan Santalo. "Platform competition: Strategic tradeoffs in platform markets." *Strategic Management Journal* 34.11 (2013): 1331-1350.

Chesbrough, H., Vanhaverbeke, W. and West, J. eds., 2014. New frontiers in open innovation. Oup Oxford.

Ching, Kenny. "Exaptation dynamics and entrepreneurial performance: evidence from the Internet video industry." *Industrial and Corporate Change* 25.1 (2016): 181-198.

Church, Jeffrey, and Neil Gandal. "Network effects, software provision, and standardization." *The Journal of Industrial Economics* (1992): 85-103.

Claussen, Jörg, Tobias Kretschmer, and Philip Mayrhofer. "The effects of rewarding user engagement: the case of facebook apps." *Information Systems Research* 24.1 (2013): 186-200.

Cusumano, Michael A., Yiorgos Mylonadis, and Richard S. Rosenbloom. "Strategic maneuvering and mass-market dynamics: The triumph of VHS over Beta." Business history review 66.1 (1992): 51-94.

Farrell, Joseph, and Garth Saloner. "Installed base and compatibility: Innovation, product preannouncements, and predation." The American economic review (1986): 940-955.

Felin, Teppo, Karim R. Lakhani, and Michael L. Tushman. "Firms, crowds, and innovation." Strategic organization 15, no. 2 (2017): 119-140.

Felin, Teppo, and Todd R. Zenger. "Closed or open innovation? Problem solving and the governance choice." Research Policy 43.5 (2014): 914-925.

Gambardella, Alfonso, Christina Raasch, and Eric von Hippel. "The user innovation paradigm: impacts on markets and welfare." Management Science 63.5 (2016): 1450-1468.

Gans, Joshua S. "Mobile application pricing." Information Economics and Policy 24, no. 1 (2012): 52-59.

Garg, Rajiv, and Rahul Telang. 2012. "Inferring app demand from publicly available data." *MIS Quarterly*.

Gawer, Annabelle, and Michael A. Cusumano. Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation. Vol. 5. Boston: Harvard Business School Press, 2002.

Ghazawneh, Ahmad, and Ola Henfridsson. "Balancing platform control and external contribution in third‐party development: the boundary resources model." *Information Systems Journal* 23.2 (2013): 173-192.

Ghose, A., Han, S. P. (2014). Estimating demand for mobile applications in the new economy. *Management Science*, *60*(6), 1470-1488.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li. "Examining the impact of ranking on consumer behavior and search engine revenue." *Management Science* 60.7 (2014): 1632-1654.

Hamilton, Barton H. 2000. "Does entrepreneurship pay? An empirical analysis of the returns to self-employment." *Journal of Political Economy* 108.3: 604-631.

Howison, James, and James D. Herbsleb. "Scientific Software Production: Incentives and Collaboration." Proceedings of the ACM 2011 conference on Computer supported cooperative work. ACM, 2011.

Hossain, T., Minor, D., & Morgan, J. (2011). Competing matchmakers: an experimental analysis. Management Science, 57(11), 1913-1925.

Jeppesen, Lars Bo, and Lars Frederiksen. "Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments." Organization science 17.1 (2006): 45-63.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?. *The Review of Economic Studies*, 76(1), 283-317.

Katz, Michael L., and Carl Shapiro. "Network externalities, competition, and compatibility." The American economic review 75.3 (1985): 424-440.

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28, no. 3 (2014): 25-48.

Kittur, Aniket, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. "The future of crowd work." In Proceedings of the 2013 conference on Computer supported cooperative work, pp. 1301-1318. ACM, 2013.

Lakhani, Karim, and Robert Wolf. 2005. Why hackers do what they do: Understanding Motivation and Effort in Free/Open Source Software Projects." J. Feller, ed. Perspect*.* *Free Open Source Software*. Cambridge, MA.

Lakhani, Karim R., and Eric Von Hippel. "How open source software works: "free" user-to-user assistance." *Research Policy* 32.6 (2003): 923-943

Langlois, Richard N. "External economies and economic progress: The case of the microcomputer industry." Business history review 66.1 (1992): 1-50.

Lee, Gunwoong, and T. Santanam Raghu. "Determinants of mobile apps' success: evidence from the App Store market." Journal of Management Information Systems 31.2 (2014): 133-170.

Lerner, Josh, Parag A. Pathak, and Jean Tirole. 2006. "The dynamics of open-source contributors." *American Economic Review*: 114-118.

Lifshitz-Assaf, Hila. (2018) "Dismantling knowledge boundaries at NASA: from problem solvers to solution seekers." *Administrative Science Quarterly*.

Lyytinen, Kalle, Youngjin Yoo, and Richard J. Boland Jr. "Digital product innovation within four classes of innovation networks." *Information Systems Journal* 26.1 (2016): 47-75.

Majchrzak, Ann, and Arvind Malhotra. "Towards an information systems perspective and research agenda on crowdsourcing for innovation." *The Journal of Strategic Information Systems* 22.4 (2013): 257

Manso, Gustavo. 2016. "Experimentation and the Returns to Entrepreneurship." *The Review of Financial Studies*, 29(9), 2319-2340.

McAfee, Andrew. and Erik Brynjolfsson, 2017. *Machine, Platform, Crowd: Harnessing our Digital Future.* W.W. Norton & Co..

Miric, Milan, Kevin Boudreau, Lars Bo Jeppesen (2019). "Protecting their Digital Assets". Research Policy.

Muller. 2009. "Gold Rush in the Smartphone Software Market." Der Spiegel Online, downloaded from http://www.spiegel.de/international/zeitgeist/the-apps-boom-gold-rush-in-the-smartphone-software-market-a-659563-2.html

Nagaraj, Abhishek. "Does copyright affect reuse? Evidence from google books and wikipedia." *Management Science* (2017).

Oestreicher-Singer, Gal, and Arun Sundararajan. "Recommendation networks and the long tail of electronic commerce." Mis quarterly (2012): 65-83.

Parker, Geoffrey G., and Marshall W. Van Alstyne. "Two-sided network effects: A theory of information product design." Management Science 51.10 (2005): 1494-1504.

Parker, Geoffrey G., Marshall W. Van Alstyne, and Sangeet Paul Choudary. *Platform Revolution: How Networked Markets Are Transforming the Economy and How to Make Them Work for You*. WW Norton & Company, 2016.

Pendharkar, Parag C. 2004. "An exploratory study of object-oriented software component size determinants and the application of regression tree forecasting models." *Information & Management* 42.1: 61-73.

Piezunka, Henning, and Linus Dahlander. "Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing." *Academy of Management Journal* 58.3 (2015): 856-880.

Rietveld, Joost, and J. P. Eggers. "Demand Heterogeneity in Platform Markets: Implications for Complementors." *Organization Science* (2018).

Roach, Michael, and Henry Sauermann. "A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry." Research Policy 39, no. 3 (2010): 422-434.

Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform competition in two-sided markets." *Journal of the European Economic Association*: 990-1029.

Roberts, Jeffrey A., Il-Horn Hann, and Sandra A. Slaughter. 2006. "Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects." *Management Science* 52.7: 984-999.

Rohlfs, J. (1974). A Theory of Interdependent Demand for a Communications Service. The Bell Journal of Economics and Management Science, 16-37.

Sauermann, Henry, and Wesley M. Cohen. 2010. "What makes them tick? Employee motives and firm innovation." *Management Science* 56.12: 2134-2153.

Sauermann, Henry, and Chiara Franzoni. "Crowd science user contribution patterns and their implications." Proceedings of the National Academy of Sciences 112.3 (2015): 679-684.

Schilling, Melissa A. "Technology success and failure in winner-take-all markets: The impact of learning orientation, timing, and network externalities." *Academy of Management Journal* 45.2 (2002): 387-398.

Seamans, Robert, and Feng. "Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers." *Management Science* 60, no. 2 (2013): 476-493.

Scott Morton, Fiona M., and Joel M. Podolny. 2002. "Love or money? The effects of owner motivation in the California wine industry." *Journal of Industrial Economics* 50.4: 431-456.

Shankar, Venkatesh, and Barry L. Bayus. "Network effects and competition: An empirical analysis of the home video game industry." *Strategic Management Journal* 24, no. 4 (2003): 375-384.

Stern, Scott. 2004. "Do scientists pay to be scientists?." *Management Science* 50.6: 835-853.

Suarez, F.F., 2004. Battles for technological dominance: an integrative framework. *Research Policy*, 33(2), pp.271-286.

Sundararajan, Arun, 2016. *The sharing economy: The end of employment and the rise of crowd-based capitalism*. MIT Press.
Terwiesch, Christian, and Yi Xu. "Innovation contests, open innovation, and multiagent problem solving." *Management Science* 54.9 (2008): 1529-1543.
Venkatraman, N. and Lee, C.H., 2004. Preferential linkage and network evolution: A conceptual model and empirical test in the US video game sector. Academy of Management Journal, 47(6), pp.876-892.
Verner, June, and Graham Tate. 1992. "A software size model." *IEEE Transactions on Software Engineering* 18.4: 265-278.
von Hippel, Eric 2017. *Free Innovation*. Cambridge MA: MIT Press.
Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. MIS quarterly, 649-676.
Waldfogel, Joel. 2014. "Digitization and the quality of new media products: the case of music." *Economic Analysis of the Digital Economy*. Eds Avi Goldfarb, Shane Greenstein, and Catherine Tucker. University of Chicago Press. 407-442.
Waldfogel, J. (2018). Digital Renaissance: What Data and Economics Tell Us about the Future of Popular Culture. Princeton University Press.
Waldfogel, Joel, and Imke Reimers. 2015. "Storming the gatekeepers: Digital disintermediation in the market for books." *Information Economics and Policy* 3: 47-58.
Weeds, H. (2012). Superstars and the long tail: the impact of technology on market structure in media industries. Information Economics and Policy, 24(1), 60-68.
West, Joel, and Marcel Bogers. "Leveraging external sources of innovation: a review of research on open innovation." *Journal of Product Innovation Management* 31.4 (2014): 814-831.
Wu, Chorng-Guang, James H. Gerlach, and Clifford E. Young. "An Empirical Analysis of Open Source Software Developers' Motivations and Continuance Intentions." *Information & Management* 44.3 (2007): 253-262.
Xu, Lei, Tingting Nian, and Luis Cabral. 2016. What makes geeks tick? A study of stack overflow careers. Working paper.
Yin, Pai-Ling, Jason P. Davis, and Yulia Muzyrya. "Entrepreneurial innovation: Killer apps in the iPhone ecosystem." American Economic Review 104, no. 5 (2014): 255-59.
Zhu, Feng, and Marco Iansiti. "Entry into platform-based markets." Strategic Management Journal 33, no. 1 (2012): 88-106.

TABLE 1 How the Analytical Framework Relates to Platform Design and Developers

COST STRUCTURE		
Notation	**Marketplace Characteristic**	**Examples of Corresponding Platform Design Choices**
w_{min}	Minimum development costs for developers	• Challenge/complexity of the development task • Definition of minimum quality standards • Availability of pre-built platform functions, frameworks (to exploit and modify) • Availability of modifiable examples • Simplicity of documentation and development tools • Extent of platform-specific investments • Access fees levied on developers
R'_w	Rate at which discretionary expenditures translate to revenues	• Availability of powerful development tools • Extent of access to platform capabilities • Limits set of development resources (e.g., maximum memory capacity, etc.)

PAYOFF STRUCTURE		
Notation	**Marketplace Characteristic**	**Examples of Corresponding Platform Design Choices**
β	Non-pecuniary payoffs	• Developer public profiles of accomplishments, skills • Learning and interaction forums • Leader boards • Socialization, events and "community" development • Degree to which development problems are intrinsically interesting, challenging, creative • Marketing/catering the platform to specific developer groups (students, users, industry associations, etc.)
R, p	Developer income expectations	• Rules or restrictions on developers charging users • Overall platform marketing, marketplace growth • All platform design affecting externalities among developers (competitive intensity or network effects) • Enforcement of developer IP rights

Notes. Here the analysis focuses on non-discriminatory conditions of platform access and marketplace entry, and therefore on costs and benefits of market participation. Deliberate regulation of access is not considered.

TABLE 2 Main Variable Definitions

Variable	Unit of Observation	Unit	Definition
NumProducts	503 subcategories	[Developer Count]	Count of products offered in a given subcategory
Games	503 Subcategories	[Indicator]	Indicator switched to 1 for each of the 58 subcategories featuring games
LowMinCost	503 Subcategories	[Indicator]	Indicator switched to 1 for subcategories for which *MinFileSize<.063*
MinFileSize	503 Subcategories	[MB]	Smallest software app, in megabytes, of all apps appearing within a given subcategory
NumDevelopers	503 Subcategories	[Developer Count]	Count of unique developer s offering products in a given subcategory
Revenues	503 Subcategories	[$000's]	Total revenues of all apps in a given subcategory
Product Category Fixed Effects	43 Categories	[Indicator]	Set of fixed effects corresponding to the set of 43 parent categories mapping to each of the 503 subcategories
FileSize	693,541 Products	[MB]	Amount of memory required to save a software app on a device
NumVersions	693,541 Products	[Count]	Number of revisions or new versions released up to the period of observation
Price	693,541 Products	[$]	Prices listed on App Store to download app
UrlSuffix	503 Subcategories	[Percent]	Number of developers with a website url address that does not end in ".com," divided by total number of developers with a website
UserRatings	693,541 Products	[5-Point Scale]	Mean of all ratings received from users of the app
UserRatingsCount	693,541 Products	[Count]	Total number of ratings received from users of the app

TABLE 3 Distribution of 503 Subcategories Across 43 Categories and *MinFileSize*

Category	Subcategories in Category	*MinFileSizes* [MB] in Subcategories within the Category		
		Mean	Min.	Max.
Adventure Games	3	0.182	0.048	0.287
Hobby Games	17	0.133	0.031	0.358
Strategy Games	7	0.128	0.020	0.348
Action Games	20	0.112	0.024	0.364
Sports Games	6	0.112	0.072	0.163
Themed	23	0.085	0.017	0.223
Animals	9	0.080	0.030	0.235
Children's Games	2	0.079	0.045	0.113
Trivia Games	3	0.069	0.042	0.100
Educational Games	4	0.068	0.042	0.121
Sports	22	0.063	0.022	0.330
Beauty Fashion & Style	7	0.062	0.024	0.115
Social Tools & Utilities	7	0.061	0.023	0.145
Puzzle Games	9	0.061	0.029	0.174
Card & Casino Games	6	0.060	0.030	0.089
Religion	9	0.058	0.025	0.189
Driving & Navigation	17	0.056	0.022	0.177
Music & Audio	29	0.054	0.020	0.196
Home Household & Garden	8	0.051	0.022	0.085
Family & Children	4	0.050	0.031	0.065
Professional Information & Services	6	0.050	0.025	0.130
Entertainment & Game Related	29	0.049	0.020	0.178
Reference & Culture	17	0.047	0.023	0.109
Weather	4	0.047	0.026	0.075
Arcade Games	5	0.047	0.032	0.064
Medical	26	0.046	0.017	0.104
Education & Learning	33	0.046	0.022	0.122
Books	8	0.045	0.032	0.107
Travel & Local	13	0.045	0.021	0.075
Healthy Food & Lifestyle	12	0.044	0.029	0.083
Office Job & Tools	34	0.044	0.019	0.161
Connecting & Tools	17	0.044	0.019	0.112
Sex & Love	7	0.044	0.019	0.089
Food & Beverages	9	0.041	0.025	0.085
Device & Developer Tools	9	0.039	0.019	0.110
Photo & Video	19	0.038	0.017	0.091
Shopping & Commerce	10	0.038	0.022	0.060
Calendars & Clocks	7	0.037	0.022	0.069
Your Daily Communication	2	0.037	0.035	0.039
Finance Payments & Insurance	11	0.037	0.021	0.077
Board Games	3	0.033	0.025	0.048
News & Information	5	0.031	0.024	0.043
Chat & Short Messaging	5	0.028	0.019	0.036
	503			

TABLE 4 Motivations of Part-Time and Full-Time Developers (Sorted by Difference)

Sources of motivation:	Part-Time	≥ 1 Workers	diff.	s.e.	
it's a hobby or personal interest outside my main job	.54	.18	.36	(.05)	***
to learn new skills	.75	.51	.24	(.04)	***
for fun	.66	.49	.17	(.05)	***
to increase my job/career prospects	.38	.24	.14	(.05)	***
to use the app myself	.54	.40	.14	(.05)	***
to be part of the app developer community	.28	.21	.07	(.04)	**
to see other people using my app creations	.60	.57	.03	(.05)	
to be creative, to create new things	.73	.73	.01	(.04)	
maybe I'll get rich	.39	.40	-.02	(.05)	
to build my reputation as a developer	.30	.34	-.05	(.04)	
to tackle especially interesting technical / development problems	.29	.34	-.05	(.04)	
to do especially challenging things	.39	.46	-.07	(.05)	
to be part of an exciting industry	.41	.53	-.13	(.05)	**
to meet interesting people	.10	.23	-.13	(.03)	***
to be an entrepreneur	.46	.61	-.15	(.05)	***
to make an income	.49	.73	-.24	(.05)	***

Notes. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations is 809 developers, in total, of which 135 are part-time developers. Values indicate fractions of respondents replying in the affirmative that this motivation is important.

TABLE 5 Nonlinear Relationship between Incremental Reductions in Minimum Development Cost and Numbers of Developers

Dep. Var.:	*NumDevelopers*				
	Simple	Piece-Wise Linear Model	Category Fixed Effects	Demand-Side Control	Supply-Side Control
Model:	(1)	(2)	(3)	(4)	(5)
MinFileSize		-159	-706*	-738	-706*
		(358)	(425)	(985)	(424)
LowMinCost *(I{MinFileSize<.063})*	494***	1,252***	1,256***	1,316***	1,256***
	(49)	(145)	(147)	(499)	(147)
MinFileSize × *LowMinCost*		-20,914***	-21,099***	-21,686***	-21,090***
		(2,963)	(3,005)	(8,148)	(3,015)
Subcategory Revenues				-.1	
				(.1)	
InternationalUrl					-.1
					(.5)
Category FE			Y	Y	Y
Constant	362***	381***			
	(27)	(60)			
Adj-R^2	.09	.16	.21	.21	.21
R^2	.09	.17	.28	.28	.28

Notes. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations is 503 app subcategories.

TABLE 6 Minimum File Size vs. Other Statistics of Subcategory File Sizes

Dep. Var.:	*NumDevelopers*					
	Min Cost Proxy	Other (Endogenous) Measures of Subcategory File Sizes				
Model:	(1)	(2)	(3)	(4)	(5)	(6)
MinFileSize	-4,291*** (614)					
1stPctlFileSize		7.4e-4** (2.9e-4)				
5thPctlFileSize			2.7e-4*** (9.9e-5)			
10thPctlFileSize				1.1e-4 (7.2e-5)		
MeanFileSize					-1.0 (0.9)	
StdDevFileSize						-4.9e-7 (4.5e-7)
Category FE	Y	Y	Y	Y	Y	Y
Coeff vs *MinFileSize* Coeff.		-.00	-.00	-.00	.00	.00
Coeff w/ FE : Coeff w/o FE	1.1	1.8*	2.0**	1.8*	0.5**	0.6*
Adj-R^2	.12	.01	.02	.01	.05	.05

Notes. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations is 503 app subcategories.

TABLE 7 Regression Results for Numbers of Highest-Quality Products

Depvar:	*Number of Highest Rated Products, >4.5 out of 5 Rating*					
	All counted, > 0 ratings	> 10 ratings	> 50 ratings	> 500 ratings	> 5000 ratings	95th pctl, > 150,644
	(1)	(2)	(3)	(4)	(5)	(6)
LowMinCost *(I{MinFileSize<.063})*	157.8***	30.6***	10.7**	2.4*	0.3*	0.03*
	(22.1)	(7.9)	(4.3)	(1.6)	(.2)	(.02)
Constant	111.1***	31.89***	15.85***	4.9***	0.6***	.01
	(17.2)	(7.5)	(4.7)	(1.7)	(.2)	(.01)
(LowMinCosts + Const) / Const	242%	196%	167%	149%	140%	437%
Adj-R^2	.14	.31	.37	.37	.30	.05

Notes. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations is 503 app subcategories.

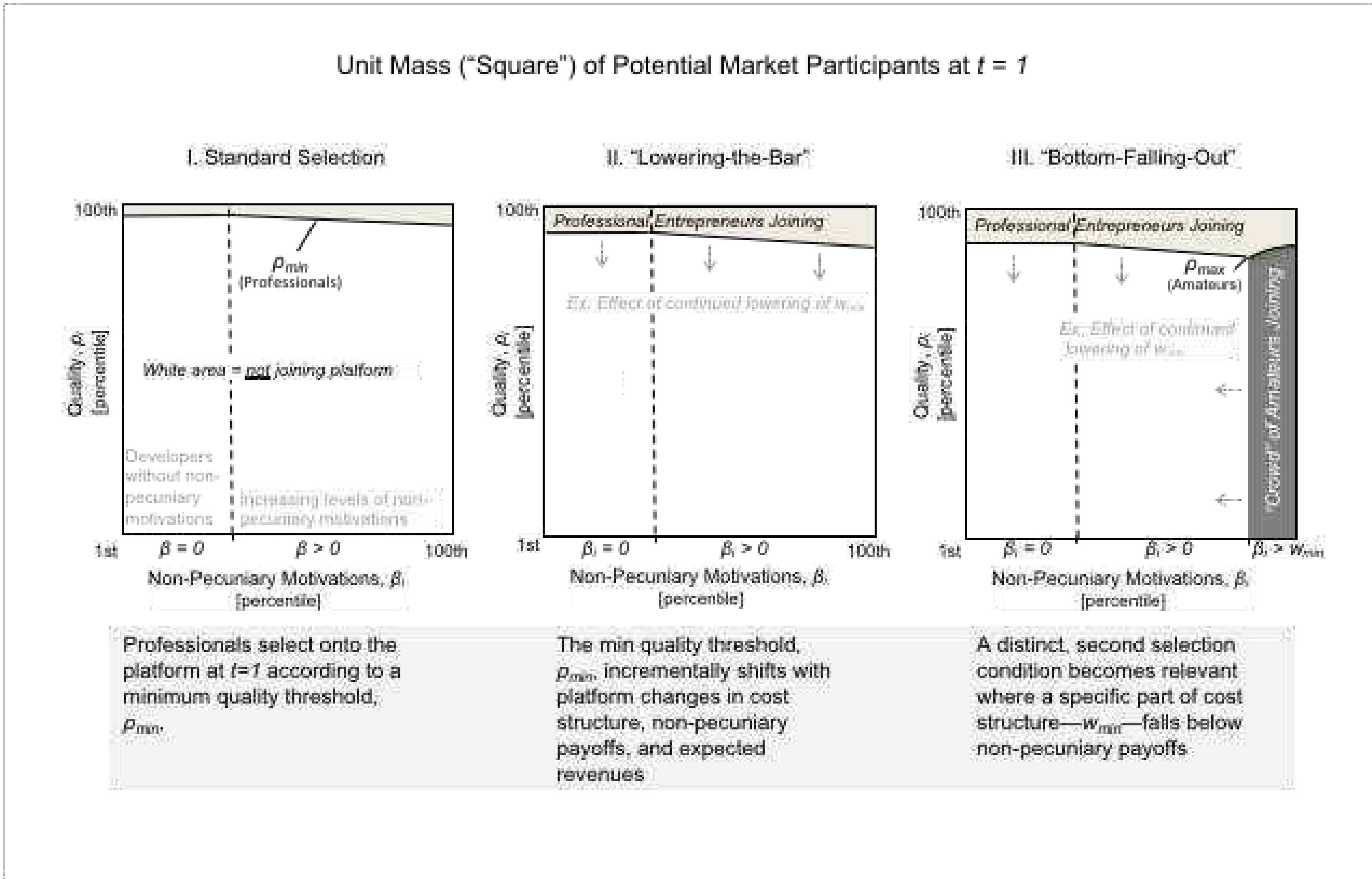

Figure 1 Figure 1 Professionals and Amateurs Join in Response to Different Thresholds and Different Selection Conditions

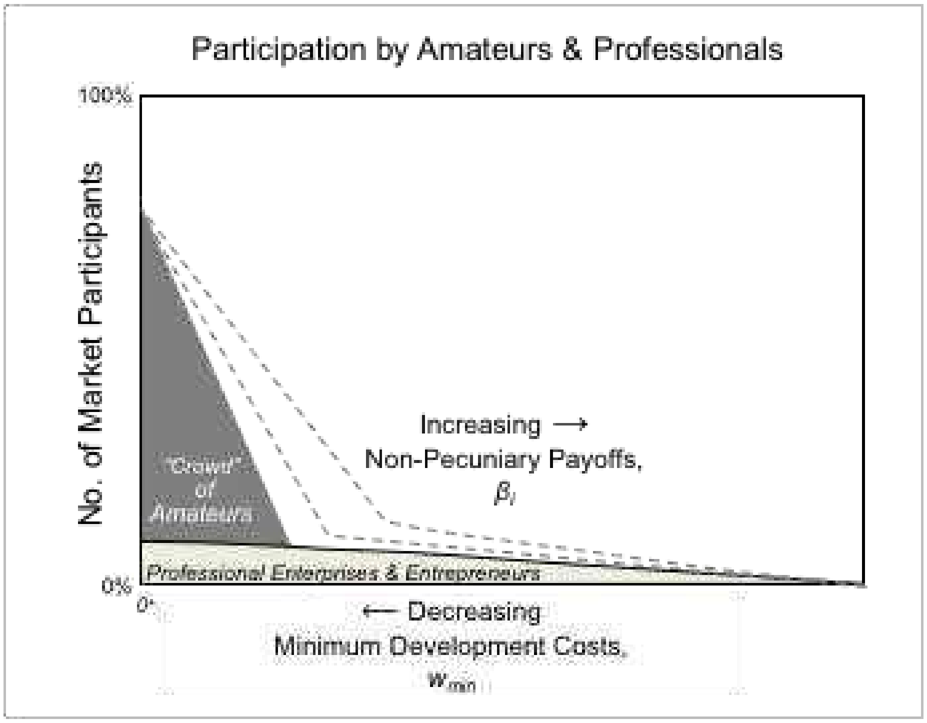

Figure 2 Summary of Main Predictions

Notes. The graph shows the total mass of complementors joining the platform. This is the sum of all professionals and amateurs joining at *t=1* and the overlapping generation that continues at *t=2*. Therefore, the 100% on the vertical axis corresponds with a mass of two.

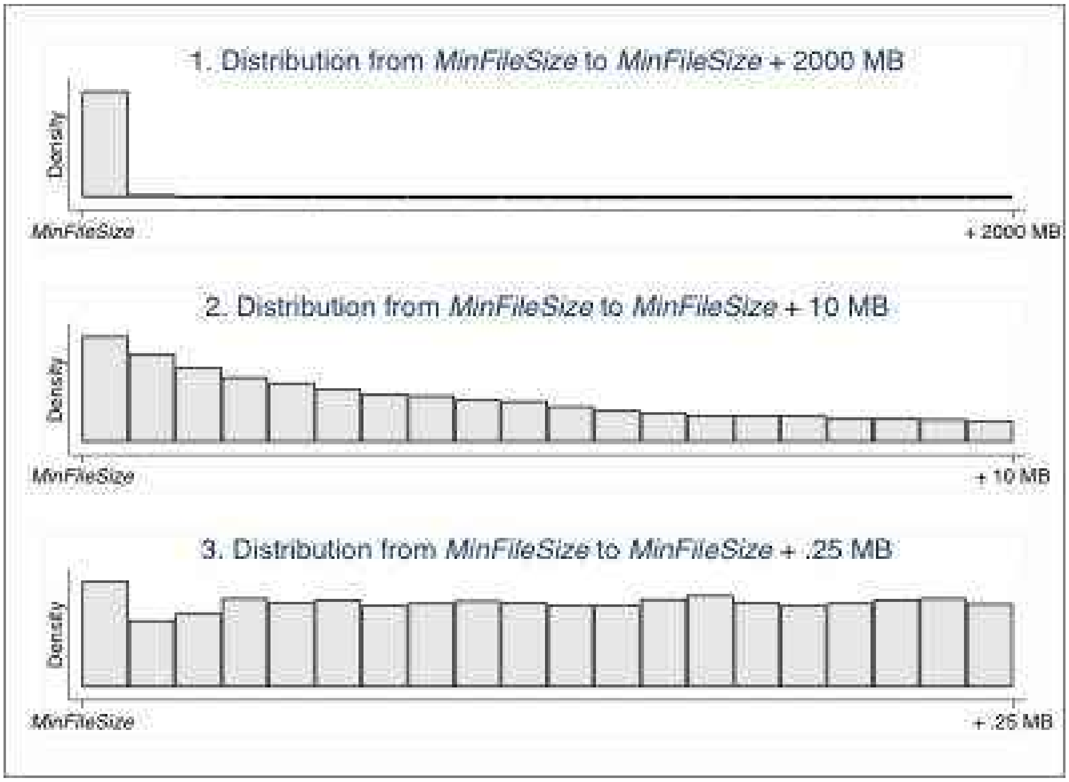

Figure 3 Within-Subcategory Product Size Distribution above *MinFileSize*:
"Thick Left Bunching" vs. "Thin Right Tail"

Notes. Number of observations is 693,541 products. Panels 2 and 3 do not show the entire domain of file sizes, but rather focus on just the lower extreme of the overall distribution, with greater "magnification." Vertical heights represent relative densities (probabilities) taken from all subcategories. The area (probability) in Panel 1 therefore sums to one, where those in Panels 2 and 3 sum to values less than one, as they censure the right hand side of the overall distribution.

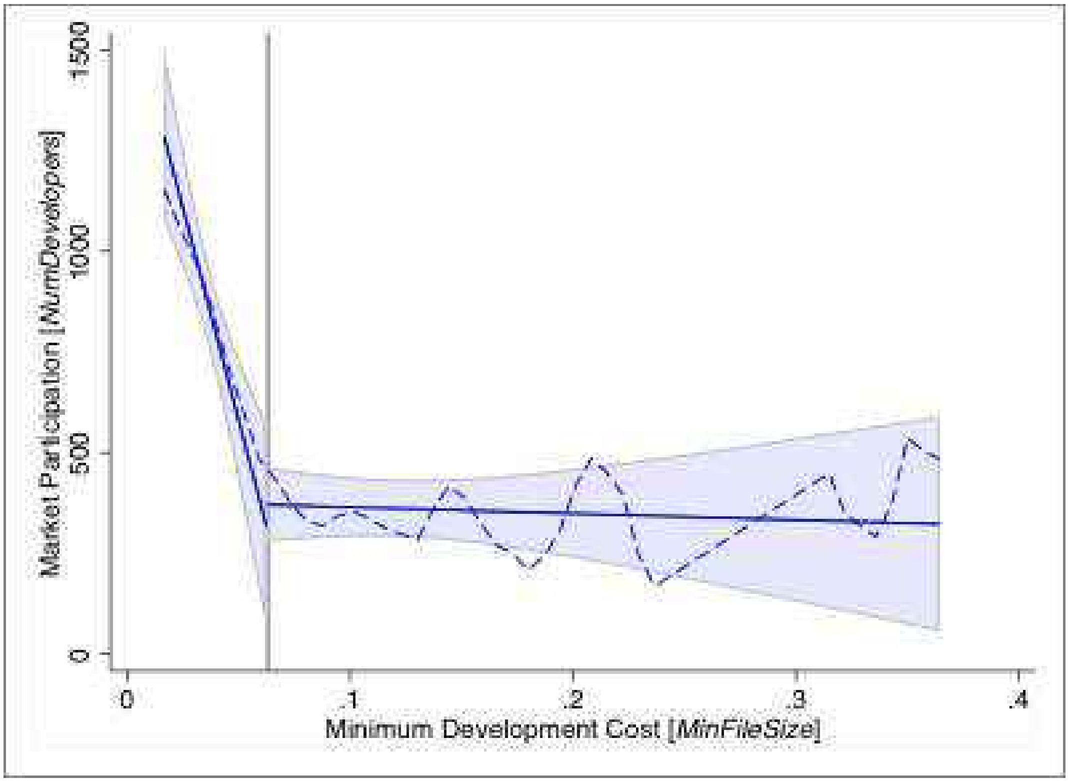

Figure 4 Discontinuous Relationship between Market Participation and Minimum Costs: Fully-Flexible Non-Parametric and Unconstrained Linear Piece-Wise Estimates

Note: The figure presents the unconstrainted piece-wise linear specification and associated 95% confidence intervals, along with a flexible non-parametric estimate using locally-weighted least-squares, weighted by second-order Epanechnikov kernel. Number of observations is 503 product subcategories.

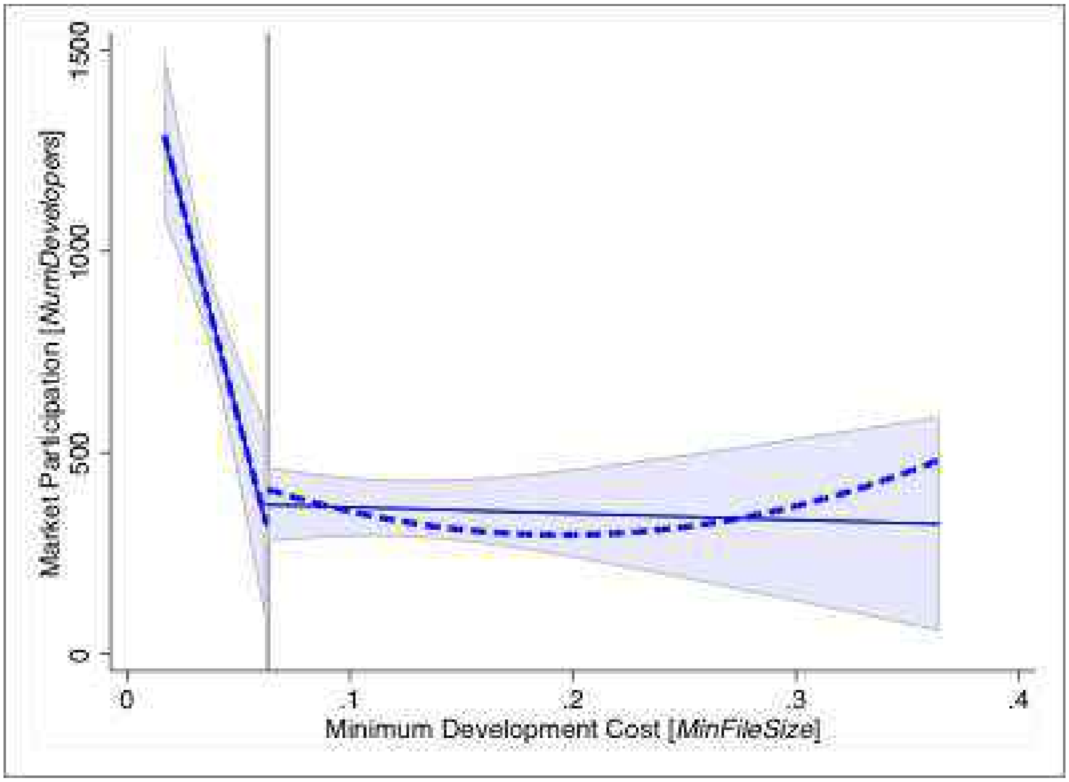

Figure 5 Piecewise Affine (Quadratic) Specification Leads to Statistically Identical Results: No Evidence of Curvilinearity on either Side of the Discontinuity

Note: The figure presents the unconstrainted piece-wise linear specification and associated 95% confidence intervals from Figure 4. Superimposed are fitted affine/quadratic models of the data—including constant, linear and quadratic terms—on either side of the earlier estimated point of discontinuity, (*MinFileSize* = 0.063). Number of observations is 503 product subcategories.

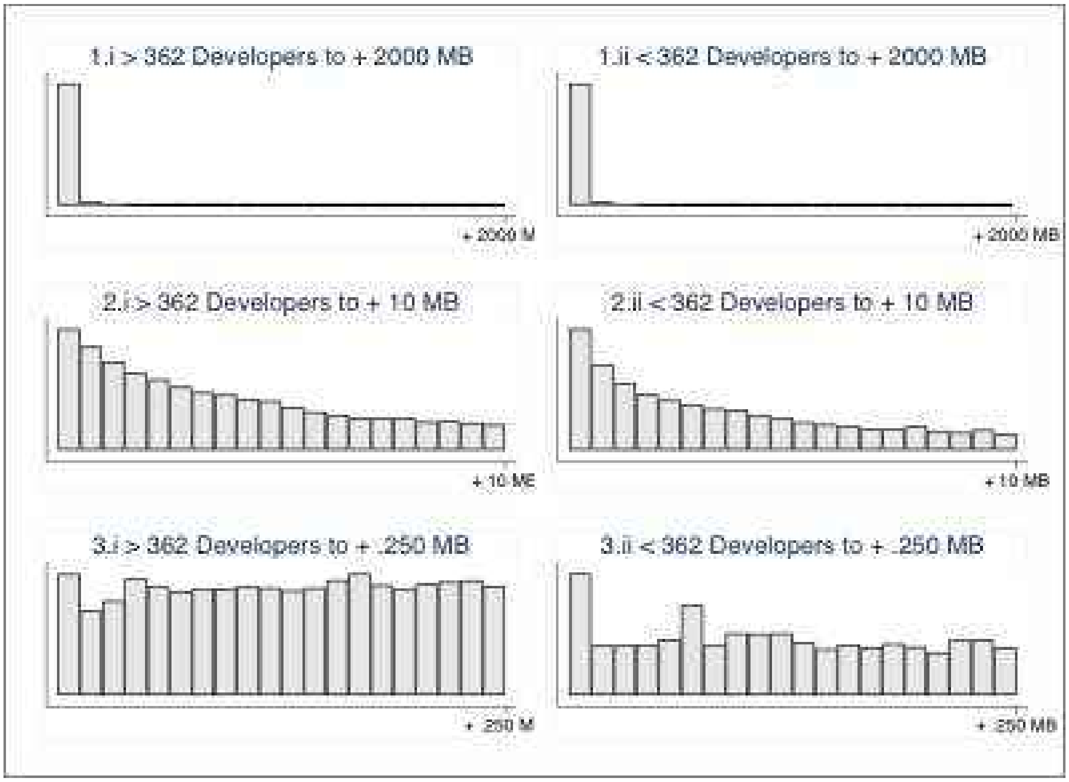

Figure 6 Within-Subcategory Product Size Distribution above *MinFileSize*: Comparing Subcategories with Relatively Many or Fewer Developers

Notes. Patterns presented above are similar to those presented in earlier Figure 3, but stratified by cases with relatively many or few developers. The cut-off between many and few is defined as 362, as this is the mean number of developers in cases to the right of the point of discontinuity in Figure 4. Number of observations is 693,541 products. Note, as in Figure 3, Panels 2 and 3 focus on just the lower or leftward part of the distribution.

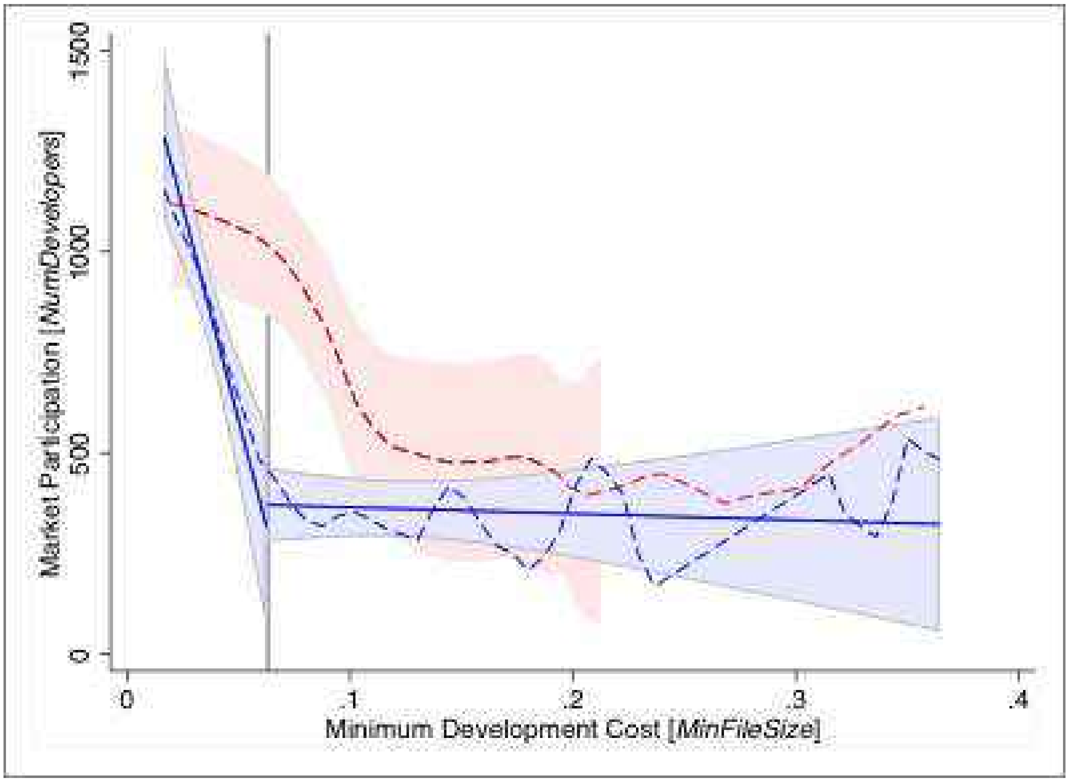

Figure 7 Discontinuity Shifts to the Right for Subcategories with Higher Non-Pecuniary Motivations (Games, in red dash)

Note: Same as Figure 4, with flexible non-parametric estimate superimposed for games-related subcategories. Number of games observations is 58 product subcategories. Confidence intervals are not reported over the entire domain for games, as standard errors grow too large to show within the range presented.

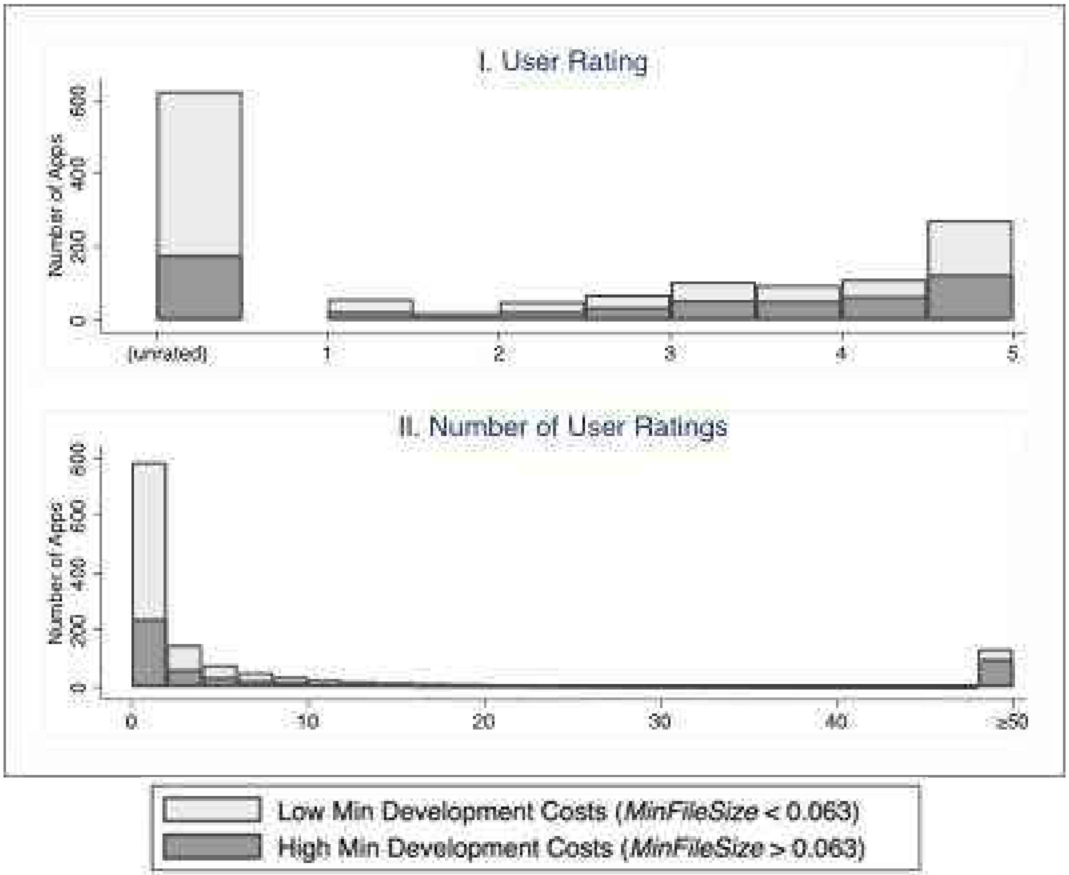

Figure 8 Distribution of Products per Subcategory, Stratified by Minimum Development Costs

Notes. The histograms present mean frequency of products at different increments per subcategory, based on the 693,541 product observations. Histograms are stratified by subcategories above or below the point of discontinuity.

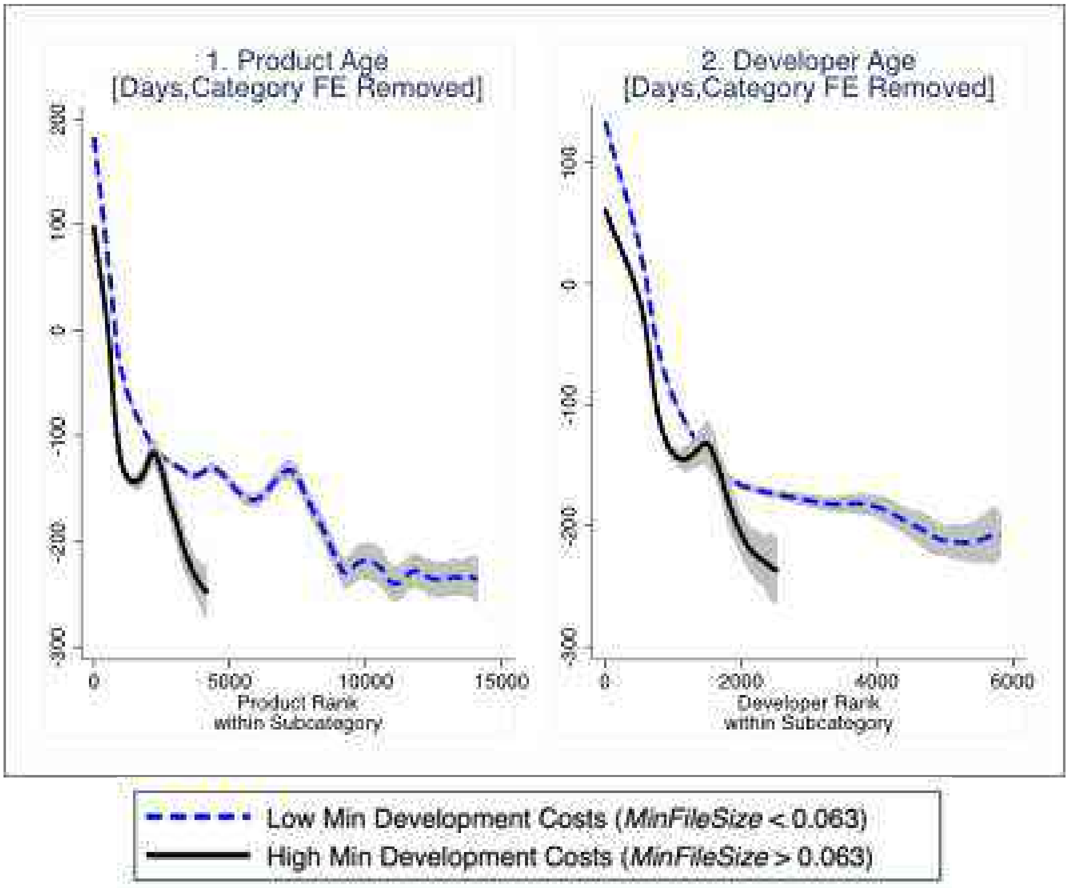

Figure 9 Persistence on the Platform with Lower versus of Product Life by Product Rank within Subcategory, Above and Below the Discontinuity (Blue)

Note. The analysis draws comparisons from subcategories on either side of the point of discontinuity (*MinFileSize* = 0.063), at each rank within their respective subcategories, controlling for 43 category fixed effects. The flexible nonlinear relationships are estimated with locally-weighted least-squares , weighted by second-order Epanechnikov kernel. Number of observations in Panel 1 and is 693,541 products; Panel 2 is 192,372 developers.

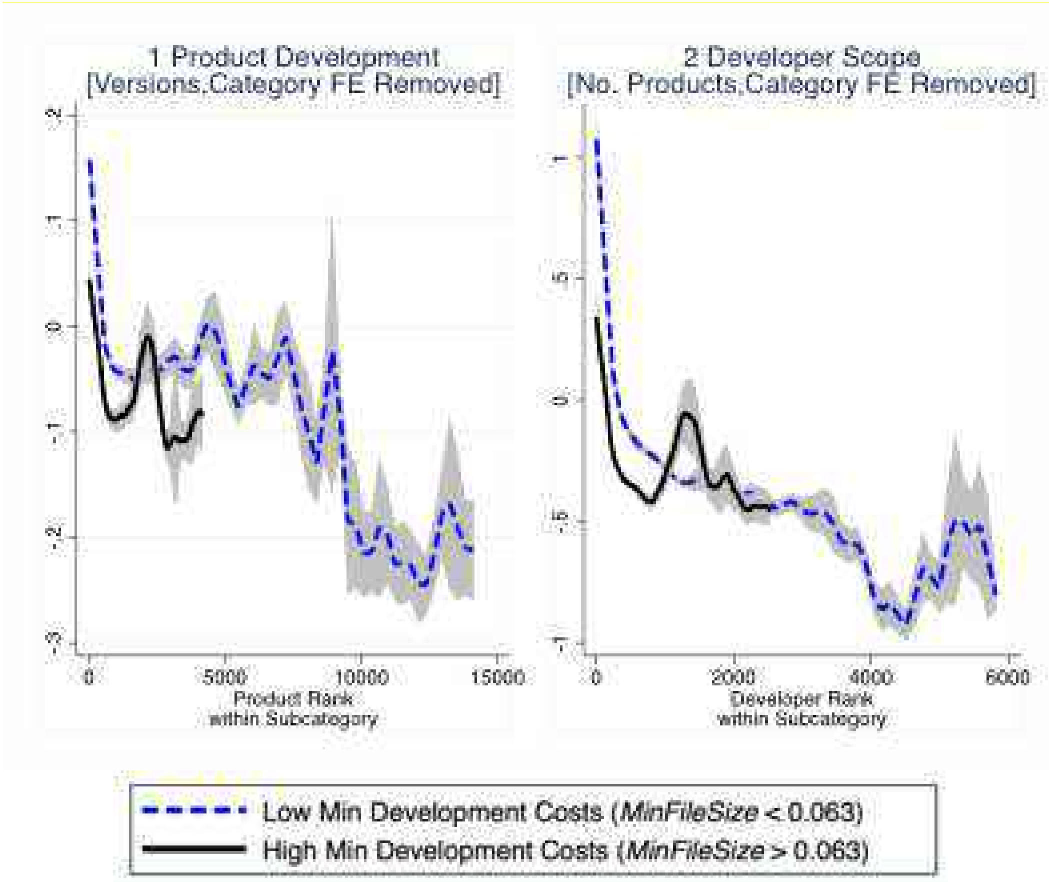

Figure 10 Product Development Activity by Developers at Different Ranks in their Subcategories, Above and Below the Discontinuity (Blue)

Note. The analysis draws comparisons from subcategories on either side of the point of discontinuity (*MinFileSize* = 0.063), at each rank within their respective subcategories, controlling for 43 category fixed effects. The flexible nonlinear relationships are estimated with locally-weighted least-squares , weighted by second-order Epanechnikov kernel. Number of observations in Panel 1 and is 693,541 products; Panel 2 is 192,372 developers.