Amateurs

Kevin J. Boudreau*

April 7, 2018

Abstract

Platforms often have "crowds" of amateurs working on them as complementors, in other cases professional entrepreneurs—or both. What can a platform owner do to implement these outcomes? I document evidence on mobile app developers showing that just small, incremental changes in platform design—related to the bare minimum costs required to build an app and factors affecting non-pecuniary payoffs—can lead the "bottom-to-fall-out" of the market to amateurs. Where the bottom-falls-out, there is a flood of lowest-quality developers who nonetheless are long-lived on the platform and engage in relatively high development activity. I find no evidence that amateurs crowd-out development activity of top developers in this context. Moreover, the bottom-falling-out is associated with the generation of significantly greater numbers of highest-quality products. I discuss several interpretations.

Keywords: Platforms, complementors, crowds, long tail, apps, minimum viable products.

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

* Northeastern University and NBER. kevin@neu.edu Early draft. Comments welcome. I would like to acknowledge helpful comments from seminar participants at Copenhagen Business School, Harvard University, London School of Economics, IESE, the NBER Summer Institute, New York University, University College London, University of Southern California, University of Michigan, Warwick University, the International Industrial Organization Conference. The paper was also influenced by helpful comments from Carliss Baldwin, Dan Elfenbein, Shane Greenstein, Brent Hickman, Nicola Lacetera, Karim Lakhani, Gwen Lee, Hong Luo, Jenny Kuan, Justin Johnson, Alan MacCormack, Dylan Minor, Ramana Nanda, Imke Reimers, Chris Stanton, Steve Tadelis, Thales Teixeira, Neil Thompson, Eric von Hippel, and Joel Waldfogel. This work also benefitted from collaboration in related topics with Lars Bo Jeppesen (CBS) and Milan Miric (USC). Excellent research assistance was provided by Richard Paulsen and Gourishankar Seal. Executives at Priori Data contributed proprietary data to this study. I acknowledge research funding from Harvard Business School Northeastern University for this study. All errors are my own.

Electronic copy available at: https://ssrn.com/abstract=2988308

1 Introduction

Rather than always using internal R&D or outsourcing to innovation partners, firms are increasingly platforming to external large crowds—organizing them using a variety of approaches in a wide range of fields from data science, to media and graphical design, natural sciences, fashion and finance and beyond. Important scholarly advances in understanding this growing use of crowds now come from multiple research traditions.[1]

It is now increasingly common to see large crowds of amateurs developing and innovating alongside professional complementors in instances of *platform-based marketplaces* (Bakos, 2001; Krishnan and Gupta, 2001; Gawer and Cusumano, 2002; Rochet and Tirole, 2003; Varian, 2010; Ghazawneh and Henfridsson, 2013; Parker et al., 2016). For example, shortly after the launch of the Apple AppStore, 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 a recurring $100 annual platform access fee and on-going capital expenses and other opportunity costs, a large bulk of developers persist in on-going development and offering products on the AppStore without any reasonable expectation of earning revenues—much less an expectation of a positive income, even one below regular competitive returns. (Modal product revenues across products on the AppStore are zero.) A similar mix of professionals and amateurs can be seen on many other well-known platforms such Youtube, Shutterstock, Podcasts, Etsy, and Kindle Self-Publishing. In other cases, such as merchants on Amazon.com, we observe mostly professional enterprises. In still other cases, complementors are dominated by crowds of amateurs, as in the makers of "add-ons" for Mozilla's Firefox web browser. *Under what conditions do crowds of amateurs and professional complementors join on a platform? Can this be influenced by platform design?* How are amateurs and professionals really fundamentally different?

Past research stresses the importance of having large *numbers* of complementors on a platform-based marketplace (e.g., Church and Gandal, 1992; Katz and Shapiro, 1994; Gawer and Cusumano, 2002; Schilling, 2002; Shankar and Bayus, 2003; Suarez, 2004; Parker and Van Alstyne, 2005, 2017; Brynjolfsson, et al., 2003, 2011; Cennamo and Santalo, 2013; Bresnahan and Greenstein, 2014). Less research has investigated the *heterogeneity* of complementors[2]—and particularly to determinants and implications of amateurs and professionals. Making progress on amateurs, in particular, might be a useful building block

[1] See, for example, von Hippel, 2005; Benkler, 2006; Terwiesch and Xu, 2009; Baldwin and von Hippel, 2011; Kohler et al., 2011; Zhang and Zhu, 2011; Afuah and Tucci, 2012; Murray et al. 2012; Majchrzak and Malhotra, 2013; Muchnik et al., 2013; Chesbrough and Brunswicker, 2014; Felin and Zenger, 2014; Kittur et al., 2014; Mollick and Nanda, 2015; Piezunka and Dahlander, 2015; Sauermann and Franzoni, 2015; Chen and Horton, 2016; Glaeser, et al. 2016; Su, et al., 2016; Sundararajan, 2016; Felin, et al. 2017; Lakhani, 2017; McAfee and Brynjolfsson, 2017; Nagaraj, 2017; Powell, 2017; Lifshitz-Assaf 2018.

[2] See Rietveld and Eggers (2018) for an exception, examining how changes early adopters to late adopter users over time affects success of different complementors.

Electronic copy available at: https://ssrn.com/abstract=2988308

in better understanding platform positioning, performance, innovativeness and efficiency[3] (a question I return to after presenting main results). It also appears that this mix of amateurs and professionals might be influenced, inasmuch as we see different mixes of amateurs and professionals on platforms in the same industry, such as the corporate and user-generated video-sharing platforms (e.g., Ching, 2016). (Addressing this questions, as will be seen, is also an opportunity to draw together several important strands of previously separate research streams.)

The consensus dictionary definition of amateurs—those engaging in a pursuit on an *unpaid* basis—provides a useful start. I proceed with a yet weaker working definition: those who do not *require* payment to engage in a pursuit. This definition is attractive in neither presupposing skills nor eventual outcomes of amateurs.[4] Further, this definition helpfully distinguishes hobbyists, learners, and fanatics (willing to participate and incur personal expenses without any reasonable prospect or expectation of positive income) from cases in which workers and entrepreneurs are willing to accept lower-than-regular income, but nonetheless expect some form of payment (e.g., Hamilton, 2000; Scott Morton and Podolny, 2002; Stern, 2004; Åstebro et al., 2014; Sauermann and Cohen, 2010).

I develop empirical hypotheses by integrating existing standard theories of market entry and entrepreneurial selection (e.g., Bresnahan and Reiss, 1991; Hamilton, 2000; Manso, 2016) into a single compact framework. Rather than presume the existence of distinct groups of amateurs and professionals (cf. Gambardella et al. 2016), instead I summarize the range of hackers, students, hobbyists, funded startups, tinkers, learners, user innovators and other possible complementors as coming from some continuous distribution in "quality" (ex-ante expected ability to generate income) and non-pecuniary motivations. In this latter regard, the framework builds on research demonstrating multiple sources of motivations among online and platform-based developers apart from just income, including an interest in creative or challenging problem-solving, learning, building reputation, social interaction, expanding one's network, and user innovation motives. (See, for example, important contributions across a range of institutional approaches: 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; Zhang and Zhu, 2011; Von Krogh et al., 2012; Agrawal, 2014; Lei, et al., 2016; Xu, 2016; von Hippel 2017.) Apart from different types, the framework allows for differences in choices in product development expenditures and long-livedness on the platform.

Crucially, the the framework leads to two distinct solutions or sets of conditions under which complementors join a platform. A first (standard) solution shows that professionals' join according to a *minimum quality threshold*. Further, the minimum threshold incrementally shifts ("raising- or lowering-the-bar")—affecting marginal or "fringe" professionals—by altering platform design to change complementors' cost structure (e.g., provision of development tools, documentation, high powered APIs, platform access fees), levels of non-pecuniary payoffs (e.g., public profiles of complementor accomplishments, social interaction forums), or complementors' income potential (e.g., contractual proscription from charging

[3]To the extent a platform can support large numbers of amateur complementors without creating undue crowding or congestion (e.g., Bresnahan et al. 2014), findings across several literatures hint at possible benefits of amateurs on platform outcomes (e.g., von Hippel 1998**,** Jeppesen and Frederiksen, 2006; Jeppesen and Lakhani, 2010; Kittur et al. 2013; West and Bogers, 2014; Waldfogel and Reimers, 2015; Aguiar and Waldfogel, 2016, 2018; Lyytinen, et al. 2016; Sauermann and Franzoni 2015).

[4]Consider, for example, unpaid amateur Olympic competitors, open source software contributors, and student competitors in case competitions each might vary widely in skill and eventual outcomes.

platform users, platform marketing). See, for example, Ceccagnoli et al. (2012), Claussen et al. (2013), and Ghose et al. (2014).

A second solution or set of conditions for joining relates to amateurs. This second solution exists only where a specific component of complementors' cost structure—the *cost necessary to develop a bare minimum viable product* for complementors—falls below levels of non-pecuniary payoffs for some part of the distribution of complementors. Unlike the earlier condition for professionals, this second solution implies no minimum quality threshold. It is in this sense that the "bottom-falls-out" of the market to amateurs.

Complementors meeting this second set of conditions are predicted to be relatively long-lived on the platform and thus able to accumulate in large numbers, despite tending to be of lower quality. They should also should be expected to differ in product development choices inasmuch as they are not bound by usual market discipline, and payoffs are weighted towards a non-pecuniary ones. It turns out to be an outcome of the analysis (rather than a pre-baked assumption) that this second solution implies a distinct set of complementors from professionals who not require income to be willing to participate.

With abundant prior research on entry and market participation by professionals entrepreneurs and firms, the empirical analysis focuses greatest attention on conditions under which the bottom-falls-out to amateurs. The ideal empirical test involves comparing multiple identical platforms under same conditions, while independently varying minimum development costs and non-pecuniary payoffs. Such an ideal experimental is impracticable if only because platform industries tend to consolidate around few differentiated platforms and lack of controlled conditions. To make relevant controlled comparisons, I instead exploit within-platform variation across 503 narrow subcategories—platform submarkets—on the Apple AppStore. Each submarket is a microcosm platform-based market to which the earlier predictions equally apply. Within each submarket, there are hundreds or thousands of developers and apps. For example, the data studied here include 3,759 weather apps, 528 hockey games, 273 heart monitors, 543 hypnosis apps, 1,643 games on fishing, and 1,543 Sudoku games. These 503 subcategories are associated with 43 categories, thus making it possible to compare say "arcade ball games" to "classic arcade" and other subcategories, while controlling for the wider category of "arcade games." Similarly, "call managers and tools," "download managers," and other such subcategories can be compared, while controlling for the broader category of "device and developer tools." Importantly, the details of the institutional context of the AppStore provides a rare opportunity to operationalize and measurement key concepts. As platform design is fixed, research design exploits differences across subcategories to capture variation in minimum costs and non-pecuniary payoffs.

I find empirical patterns to each be consistent with the precise theoretical predictions. Small incremental shifts in minimum development costs, once low enough, produce abrupt non-linear changes in link between numbers of developers and minimum development costs. The implications of these incremental cost changes are anything but incremental: developer numbers more than double in submarkets where the bottom has fallen out. The accompanying flood of products is dominated not just by low quality, but by very lowest-quality, by all available measures. Nonetheless, developers in these submarkets live for longer spells and make relatively high numbers of products.

In a secondary exploration of implications for the basket of complementary goods, I find no evidence that added amateurs result in any congestion or crowding (beyond what might have already been the case

with hundreds of professional developers), in terms of revenues, price, or investments in new versions (cf. Boudreau, 2012; Waldfogel, 2014; Bresnahan et al. 2014). Instead, the bottom-falling-out is associated with significantly greater numbers of high-quality complementary goods. I discuss a number of possible explanations.

The paper proceeds as follows. Section 2 develops main empirical hypotheses related how platform design influences participation on a platform. Section 3 develops describes the data used for the empirical analysis. Section 4 presents main results related to platform participation. Section 5 considers implications of findings for platform performance and the basket of products that becomes available in cases where the bottom-falls-out to amateurs. Section 6 concludes.

2 Hypothesis Development

This section develops empirical hypotheses on the basis of a simple organizing framework. Time proceeds in discrete periods. In each period, a new unit mass of potential complementors arrives (e.g., completing academic training, leaving other employment or leisure, etc.). Each complementor lives two periods or phases, $t\epsilon\{1,2\}$. Therefore, platform participation at any time is the sum of phase 1 and phase 2 complementors.

In the first "selection" phase, complementors choose whether to join the platform. Product development is uncertain and fraction $p\epsilon[0,1)$ are revealed to be successful in the sense of earning non-zero revenues (net of any variable costs), $R_{i,1}$ and $R_{i,2}$. Complementors also enjoy non-pecuniary payoffs, $\beta_i\epsilon\mathbb{R}^+$, when participating on the platform. Subscripts i and t index complementors and phase. Those joining in the first phase will continue in the second "retention" phase, if successful.

Therefore, the *selection condition* in the first phase is as follows:[5]

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

Note, the terms R, W, and p should each now be interpreted as values determined in competitive equilibrium. Further, the expression presumes that exit costs are sufficiently low that complementors return to outside options if they are unsuccessful (e.g., Manso, 2016).

The *retention condition* in the second phase is as follows:

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

Complementors differ in quality, $\rho_i\epsilon\mathbb{R}^+$, and non-pecuniary motivations, $\beta_i\epsilon\mathbb{R}^+$. Quality ρ_i together with the choice of level of opportunity costs devoted to product development $W_{i,t}\epsilon\mathbb{R}^+$ determine revenues, $R(W_{i,t},\rho_i)\geq 0$.[6] 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., $W_{i,t}=w_{min}+w_{i,t}$. (It follows that first-phase developers with probability of $1-p$ of failing invest

[5] 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.

[6] Revenues might also be determined by the numbers and types of other complementor choosing to participate and perhaps their investments in product development $W_{i,t}$, if there are congestion or network effects acting among complementors.

less than do successful second-phase developers who who invest without this uncertainty and therefore have incentives to "scale-up" as they operate.)

Within this general set-up, there are a number of dimensions of platform design that might be altered, as below.

<TABLE 1>

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

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

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

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

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, $\pi'(\rho_i)>0$. Complementors 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 complementors 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 this condition holds so long as the complementor chose to join in the first phase and was successful:

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

It follows that just fraction p of complementors from $t=1$ are retained at $t=2$. The total total number of professionals is just $1+p$ times those selecting in the first phase.

Following usual intuitions and standard results across a range of literatures, participation by professional entrepreneurs and enterprises is a matter of meeting a minimum quality threshold expression 4. Changes to each dimension of platform design listed in Table 1 will *incrementally* move this threshold up or down.[7] 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 to greater numbers of lower-quality marginal or "fringe" professionals. Panels I and II of Figure 1 illustrate the minimum quality threshold and an example of lowering the minimum quality threshold.

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:

$$w_{min}\leq\beta_{i}. \tag{6}$$

[7]Particulars depend on functional form assumptions.

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. But, the right hand side is at most zero, where $w_{min} \leq \beta_i$.

This solution follows the simple intuition that amateurs will be willing to join without any necessary expectation of revenues, so long as their non-pecuniary motivations on their own exceed opportunity costs, $2\beta_i \geq W_{i,1} + W_{i,2}$ (cf. expression 3). For the "marginal amateur," without even any incentives to make discretionary investments, $W_{i,1} = W_{i,2} = w_{min}$, or $w_{min} \leq \beta_i$.

This solution differs more radically still for retention even than for selection. Whereas only fraction p of professionals are retained; in this solution, all amateurs persist, so long as $w_{min} \leq \beta_i$ remains true. Therefore, total amateurs at any point in time is $2\times$ the number joining in the first phase (not just $1 + p$ this number).

Distinguishing Amateurs from Professionals. That there are two solutions does not, on its own, immediately imply two distinct subsets of complementors. Here, however, the two solutions are non-overlapping and immediately adjacent, in terms of the parts of the ρ and β distribution. And, although this means that professionals and amateurs are part of a continuum in their "types," the discussion here will clarify why they are discretely different in their behavior and choices.

Consider first that the $\pi^{-1}(-)$ term in the minimum quality threshold for professionals (4) is mathematically undefined for negative arguments, i.e., $w_{min} \leq \beta_i$.[8] But, this is indeed *precisely* the threshold at which the marginal entrepreneur joins. That professionals and amateurs are non-overlapping and immediately adjacent follows the observation that professionals appear to, at once, be distinct from hobbyists, learners, hackers and tinkerers—but at the same time part of a "blurry" continuum with them (see Introduction).

Consider a thought experiment. 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 fixed access fees for complementors and the bottom-falls-out, $w_{min} \leq \beta_i$, for complementors with high β's, as in Panel III of Figure 1.

Now, those for whom $w_{min} \leq \beta_i$ includes both professionals (who would have joined) even without a fee change, and now a potentially much larger group of complementors who who have earlier been below the minimum threshold (Panel II versus III of Figure 1). Therefore, there will be both amateurs and professionals where $w_{min} \leq \beta_i$. To distinguish the one from the other, we can delineate those whose endogenously chosen levels of product development expenditures do not exceed their non-pecuniary payoffs, $w_{min} + w_{i,t} < \beta_i$. Thus, while amateurs do not have a quality floor, they have a ceiling:,

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

<FIGURE 1>

[8] I.e. professionals simply do not join for negative income.

2.1 Main Predictions

Conditions Leading the Bottom-to-Fall-Out to Amateurs. Whereas professionals' minimum quality threshold shifts incrementally in relation to many platforms design choices; amateurs are affected only by the interplay of bare minimum development costs, w_{min}, and non-pecuniary payoffs, β, with the effect of leading the bottom to fall out.

As illustrated in Figure 2, general changes in minimum cost structure should initially lead to the incremental addition of additional, progressively lower-quality fringe professional complementors, until finally condition 6 is met and the bottom falls out. At this point, we will see a nonlinear increase in numbers of complementors. Platform design changes that boost non-pecuniary payoffs will lead the point at which the bottom falls out to shift to the right. Thus, a boost in non-pecuniary motivations will make it easier for the bottom-to-fall-out.

<FIGURE 2>

Differences between Amateurs and Professionals. The analysis also implies several differences—some continuous, some discrete—distinguishing amateurs from professionals:

- *Differences in Quality.* Quality of amateurs and professionals will be on a continuum. Highest-quality amateurs will be of comparable quality to lowest-quality "fringe" professionals; lowest-quality amateurs are unbounded from below. (Quality here is ex-ante expected revenues.)
- *Differences in Long-livedness and Tendency to Accumulate on the Platform.* Despite their lower-quality, developers are long-lived. They are retained inasmuch as condition (6) remains true.
- *Differences in Complementary Goods Development.* Amateurs may (somehow) make different sort of product development choices inasmuch as they are not bound by market discipline and a necessity of seeking (some non-zero level of) income. They will also disproportionately receive and be motivated by non-pecuniary payoffs. Precisely how is an empirical question.

3 Data Set

A first data source used here is a machine-collected cross-section of the 192,372 app developers supplying 693,541 apps in mid-2013. This is a period of stable growth, five years after the launch of the AppStore. This data source includes app titles and developer names, price, sales rank, version number, file size, and user ratings. Focusing on data from 2013 reflect the data collection period, but also has the advantage of allowing app-level revenues to be estimated using a procedure described by Garg and Telang (2012). The app rankings, originally based on sales and download ranks and essential to this methodology, have since been altered.

These data were matched to a machine learning-based breakdown of apps into precise subcategories, generated by Priori Data. 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. These 503 subcategories subdivide 43 larger categories. Main variables are defined in Table 2.

<TABLE 2>

A third data source, described further below, is a unique survey data set, which confirms the prevalence of a range of non-pecuniary motivations and allows differences different app categories to be detected.

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

Generally speaking, we should not consider the size of a piece of software (lines of code, memory size, etc.) to be a clear or meaningful indication of its cost or complexity. Software size varies, for example, by architectural approach, elegance and parsimony, use of external libraries, compatibility choices and other factors (e.g., Pendharkar, 2004; Verner and Tate, 1992). However, (i) the specific features of the institutional context, (ii) the focus here on bare *minimum* development costs, and (iii) the focus on differences *across subcategories* allow minimum file sizes to serve as a meaningful proxy for variation in minimum development costs, w_{min}. As shown in Figure 3, these differences across subcategories are tiny, many orders of magnitude smaller than differences in average file sizes across subcategories. (The analysis will later show these differences to reflect exogenous differences inherent to the app type.)

<FIGURE 3>

First, minimum sized apps are meaningful measures of apps of each type. For an app to be admitted to the platform, it must pass a rigorous certification program. This assures proper functioning of the app and that the app does what is claimed. This includes the app not jeopardizing the security and proper operation of the system. This includes such things as long load times, improper use of file systems and storage, or system crashing. More specific checks include violations in content and use of apps such as misuse of trademarks and logos. Any app-specific applications programming interfaces that are undocumented will lead an app to be rejected. For example, where applications use a user's GPS location, this must be for purposes of providing a location-based service–rather than just naked extraction of location data for purposes of advertising alone. Apps are also evaluated to ensure they do not contain unsuitable content, possible legal issues and obligations. Thus, Apple forbids use of open source code and external code libraries. More arbitrary aesthetic requirements exist, too, that lead to standardized presentation. This includes stipulations that the use of date and time "scrollers" be of identical width, and that a standard button system be used to encourage uniform appearances.[9]

Second, the smallest app need not indeed be the lowest cost app. Rather, variation in size of very smallest apps, particularly given the highly controlled and regulated code-use, should capture inherent differences in complexity and difficulty of the development task (Albrecht and Gaffney, 1983; Boehm et

[9]These points accompany a more general set of Apple's "Human Interface Guidelines." The requirements for approval have been changed and refined over time and have become more stringent over time. A recent summary is provided at this link http://blog.safedk.com/technology/ios-ats-apple-store-policy-rejected/. Especially notable were changes in 2017 intended to reduce undifferentiated apps by announcing "apps created from a commercialized template or app generation service will be rejected" while also imposing a series of principles for evaluating minimum quality and functions (http://sdtimes.com/app-store-review-guidelines-smbs/).

Beyond just this structured and rule-bound certification process, some features of the AppStore lead to further control and standardization of minimum constraints and functional requirements. For example, all apps transactions are managed centrally on the one Apple-sponsored AppStore (rather than having iOS-compatible apps traded through alternative channels). Also, unlike other leading mobile platforms, the Apple iOS operating system has not been allowed to fork to allow for multiple variants, supporting different hardware designs. Apple hardware is made only by one company–itself–and variations in iPhone hardware at any one time do not require multiple versions of software to run on each.

al. 2000). Even with the above described high degrees of control and standardization of code use, there will nonetheless be considerable variation in elegance in writing in code—and the minimum file size might reflect the most efficient approach, well-designed. This might in fact be a most meaningful measure of all in capturing differences in "hard" technical constraints to generating a certain kind of app with minimal code. (See empirical analysis for tests of this interpretation.) Many of these arguments will not apply to other parts of the distribution of file sizes, outside of the minimum file size and particularly for larger code bases, where there is much scope for any number of factors other than just complexity and challenge influencing the code base.

3.2 Measuring Differences in Non-Pecuniary Payoffs, β

Data collected by Miric, et al. (2018) on 809 app developers' representing 7,973 apps on the basis of a large-scale survey are useful to drawing out a set of useful distinctions to test earlier predictions related to non-pecuniary motivations. It is useful to first review the broad patterns in the data. Table 3 compares motivations self-reported by part-time and full-time developers, ordered by the differences in between part-time versus full-time developers. Ordering the data in this manner is useful to the extent we can interpret part-time developers to weight more towards amateurs than do full-time developers. Consistent with this interpretation part-time developers are more likely to report (in order) "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, they are more likely to indicate "to make an income" or "to be an entrepreneur

As large as this survey sample may be, it remains too small to meaningfully match to the large population data here. However, we can nonetheless exploit these survey data to draw inferences across subcategories. In particular, survey respondents in the 58 games subcategories were 13% ($s.e. = 0.05$) more likely to report engaging in development "for fun" than those developing for non-games related subcategories. The analysis will therefore simply exploit this knowledge that games elicit greater degrees of non-pecuniary motivations among their developers. (The empirical challenge will then be whether any observed differences in outcomes can be attributed to these differences in motivations rather than other plausible differences between games and non-games.)

<TABLE 3>

4 Main Results

Sections 4.1 and 4.2 test main predictions regarding conditions under which the bottom-falls-out to amateurs. Section 4.3 tests corollary predictions about the differences between amateurs and professionals.

4.1 Numbers of Developers & Minimum Development Costs

As summarized in Figure 2, earlier arguments predict that the developer of numbers, $NumDevelopers$, should initially incremental increase as minimum development costs, proxied by $MinFileSize$, are lowered (lowering-the-bar). However, at some point, we should see a nonlinear change in the relationship between numbers of developers and minimum costs, if the bottom-falls-out to amateurs.

The relationship is first estimated using a fully-flexible nonparametric estimator with locally-weighted least-squares calculated along the curve. Weights are provided by a second-order Epanechnikov kernel with bandwidth chosen according to "direct rule-of-thumb" local-linear method of Ruppert, Sheather and Wand (1995). As shown in Figure 4, moving from right to left along the x-axis, the results are consistent with cost reductions first leading to incremental increases in $NumDevelopers$. However, at still lower costs, there is a nonlinear and even "kinked" increase, consistent with the "bottom falling out" to amateurs. The shape is similar to that presented in predictions in Figure 2.

<FIGURE 6>

To estimate the location of the "kink" I re-estimate the relationship as an unconstrained piece-wise linear model, where model coefficients are estimated on two independent curves. The model also parametrically estimates the breakpoint between the two curves. The specification is as follows:

$$NumDevelopers_s = \begin{cases} \beta_0^{low} + \beta_1^{low} \cdot MinFileSize_s + \Theta_s^{low} + \varepsilon_s^{low} \; if \; MinFileSize_s \leq \delta \\ \beta_0^{high} + \beta_1^{high} \cdot MinFileSize_s + \Theta_s^{high} + \varepsilon_s^{high} \; if \; MinFileSize_s > \delta \end{cases}, \quad (8)$$

where s indexes product subcategories, Θ summarizes all subcategory-specific variation, and δ is the parameterized breakpoint. Parameters are estimated via maximum likelihood. Figure 4 graphically presents estimates of β_0's, β_1's, and δ. The break point is estimated to be $MinFileSize = 0.063$ megabytes ($s.e. = 0.023$). (In subsequent analysis, I will use this break point to distinguish the bottom-fall-out or not.[10]) Although the two segments of the linear model are estimated separately, the two linear pieces touch at the breakpoint.

Model estimates from the piece-wise model are reported in model (2) of Table 4. Consistent with predictions and earlier non-parametric results, the slope of the relationship is shallow and negative to the right of this breakpoint, at 159 developers per megabyte reduced to the minimum cost. To the left of the breakpoint, this slope is many more times larger in magnitude or 21,073 (159 plus 20,914) added developers per megabyte increase in minimum cost. Actual changes are considerably smaller than these coefficients might initially suggest, given that $MinFileSize$ and its variation are not just incremental, but tiny (Figure 3).

Model (1) provides a more direct interpretation of magnitude, estimating mean differences on either side of the breakpoint with a dummy capturing where the bottom-falls-out with lower minimum development cost, $MinFileSize < 0.063$. Where minimum development costs are relatively high, the mean number of developers is 362 ($s.e. = 27$), from the estimated constant term. The estimated coefficient on the low-cost dummy is, 494 ($s.e. = 49$), indicating the number of developers more than doubles to 856 (494 + 362), as the bottom falls out.

<TABLE 4>

[10] Strictly speaking, this is a statistical statement. We might expect exceptions on either side of the breakpoint.

4.1.1 Robustness

The above results closely conform to predictions summarized in Figure 2. However, before proceeding further, I investigate robustness of results in Table 5.

Omitted Variable Bias? One concern in interpreting results is that $MinFileSize$ could be correlated with unobserved factors that themselves shape *NumDevelopers,* biasing earlier estimates. To assess this possibility, I introduce the 43 product category dummies. These are stringent controls that surely capture a large portion of any possible supply-side and demand-side determinants of *NumDevelopers.* As reported in model (1) of Table 5 adding these stringent controls leads estimated coefficients on $LowMinCost$ and $MinFileSize \times LowMinCost$ to be almost precisely the same as in earlier estimates of model (2) of Table 4. The point estimate of the coefficient on $MinFileSize$ changes more appreciably, but not statistically so.

An additional test to attempt to more directly detecting spurious correlation with market potential is to add lagged subcategory revenues, as in model (2).[11] This again, has no statistical effect on model estimates. (The coefficient on revenues is itself zero. In simpler uncontrolled regressions, the relationship between numbers of developers and revenues is positive.) These results are consistent again with the tiny variation in *MinFileSize* and its relationship with *NumDevelopers* being orthogonal to other determinants of *NumDevelopers.*

<TABLE 5>

Reverse Causation? Another possible source of bias in earlier estimates relates to possible reverse causation, where say high levels of competition could itself cause lower investment and smaller file size. If this were the case, we might expect similar patterns for say the the first percentile in file size. To make starkest comparisons, I report simple linear relationships between *NumDevelopers* and both with *MinFileSize* and with *1stPctlFileSize*, in models (3) and (4). The relationship with the minimum file size is starkly different from that with the percentile. The relationship with *MinFileSize* is again consistent our interpretation as a shift in the hard technical constraint, and—importantly, here—inconsistent with the relationship with the endogenous first percentile (i.e., not bound by a hard constraint, being well above the minimum).

I repeat this comparison too with the mean level of file size, in model (5). The relationship with mean file size is simply zero. This null result is especially notable when considering the massive variation of the mean, and tiny variation of the minimum (Figure 3). Apart from mean file sizes possibly reflecting endogenous responses of some sort, the larger code base is more likely to reflect many more factors than just technical difficulty or levels of effort (see earlier discussion in Section 3.1).

<TABLE 3>

Endogeneity Bias via "Over-Sampling" of Minimum File Size? Another concern and source of bias would exist if *MinFileSize* were not to reflect hard technical constraints, but instead was just a decline of minimum file sizes with greater numbers of "draws" (an order statistic argument). To attempt to detect any such effect, model (6) replaces $MinFileSize$ with the mean value of Monte Carlo re-sampling of the

[11] Revenues are endogenous and therefore results should be interpreted with caution.

minimum file size with replacement, over 50 trials. The Monte Carlo simulated values for *MinFileSize* lead to similar coefficient estimates, as in model (6). Alternatively defining simulated measures in relation to the median (rather than mean) of simulations, or simply mechanically dropping minimum file sizes and using second-lowest file size, fails to perturb the results.

4.2 Numbers of Developers, Minimum Costs & Non-Pecuniary Motivations

As summarized in Figure 2, earlier arguments predict that a boost in non-pecuniary payoffs should shift the cost at which the bottom-falls-out "to the right," i.e., to higher levels of minimum development costs. Here, we exploit the earlier finding (Section 3.2) that those building games were 13% ($s.e. = 0.05$) more likely to report engaging in development "for fun" than those developing for non-games related subcategories. This conforms to the intuition that there are higher intrinsic motivations to develop games than for other software. This might relate to the creative process, or simply the prospect of playing the games.

The analysis here therefore compares the 58 observations related to games subcategories to those of other subcategories. As shown graphically in Figure 7, the non-linear relationship of the 58 games subcategories is indeed shifted to the right of the relationship for the wider population. (The 58 observations do not allow a break point in a piecewise model to be estimated with precision, as earlier.)

Of course, the challenge in interpreting this result is that games and non-games should surely differ in ways other than just motivations. This problem of interpretation cannot be unequivocally addressed by further measurement or testing. However, several sources of (some) assurance should be noted. First, the relevant question is not whether games and non-games are different, or even different in their numbers of developers. The narrower and more relevant question is whether they differ in their link between $NumDevelopers$-$MinFileSize$. Despite the earlier analysis finding that other determinants of *NumDevelopers* (other than *MinFileSize*) were statistically orthogonal to the relationship; it is only when we focus on this particular cut of the data do we see differences. Moreover, these are not simply differences, but rather patterns than precisely conform to quite specific predictions. (Finally, the results here might too be considered in light of the wider array of patterns predictions that will be presented in this section that precisely conform to predictions.)

<FIGURE 5>

4.3 Differences between Amateurs & Professionals

This section contrasts products and developers in subcategories where the bottom fell out (i.e., $MinFileSize < 0.063$) in relation to those in other cases.

Lowest-Quality Products. Earlier arguments predict that as the bottom falls out, added developers should not only be of lower quality, but will include those of lowest quality. Of course, it is not possible to observe ex-ante quality of developers; however, here it is possible to observe ex-post (realized) measures of product quality do exist. A most important measure of quality here are user provided ratings on a 5-point quality scale. Consistent with predictions, the vast bulk of the average 846 products added to a submarket as the bottom falls out are not just low-quality, but by very lowest quality products in terms of either having a mean rating of less than one or no rating at all, as in Panel I of Figure 8. Other possible rough

correlates of quality—the number of user ratings, file size, and numbers of versions—are each consistent with the the bulk of added developers being of very lowest quality where the bottom-falls-out.

<FIGURE 6>

Long-Livedness on the Platform. Earlier arguments also predict that that where the bottom-falls-out, developers will tend to persist in offering their products on the platform, even despite low quality. To compare long-livedness on the platform, Panel 1 of Figure 7 presents the lifespan of product on the platform, in days, from top to bottom-ranked apps—shown separately for cases where the bottom-fell-out (i.e., $MinFileSize < 0.063$) versus not. To exploit just differences across similar subcategories, the number of days is corrected (de-meaned) by product categories.costs. Rank order of an app within a subcategory is approximated by the numbers of user ratings.[12]

As might be expected, highest-ranked apps—to the left of the figure—have longest lifespan since they were first introduced. This is consistent with the very top, most successful apps being less subject to selection, and their developers having incentives for on-going development and versioning (see Panel III of Figure 7). Consistent with the main question here, the patterns are also consistent with products being longer-lived in subcategories where the bottom falls out, as the estimate curve for subcategories where the bottom-fell-out (i.e., $MinFileSize < 0.063$) is statistically higher.

Differences in Product Development. Earlier arguments predict that amateurs should somehow differ in their product development choices, inasmuch as they are not bound to regular market discipline and payoffs skew to non-pecuniary motivations. Here, I test for material differences in key product development outputs: numbers of products generated by a given developer and versioning of products, in Panels II and III of Figure 7. The most relevant question here is whether somehow the "long tail" in cases where the bottom falls out is somehow engaging in product development differently from what would be expected of the rightward tail of just fringe professionals.

As seen in Panel III of Figure 7, the numbers of versions released (corrected by category) in relation to rank order is similar for overlapping ranks in subcategories where the bottom-fell-out or not, with highest numbers of products released by highest-ranked products, and declining numbers for ever lower ranks. There is then perhaps some suggestion of a slightly higher mean number of versions created, as one proceed to the very rightward "long tail" in app types where the bottom-fell-out.

As seen in the case of Panel II of Figure 7, among very top-ranked developers, there is a similar pattern of highest-ranked developers producing greatest numbers of products, with a quite rapid decline outside of very top ranks. In the lower tail, however, the mean number of products where the bottom-fell-out is generally higher than for fringe professionals in categories where the bottom has not fallen out.

<FIGURE 7>

5 Exploration of Implications of the Bottom-Falling-Out

Following the earlier main analysis, this section provides a preliminary exploration of how the bottom-falling-out affects the basket of complementary goods. The number of apps within each subcategory

[12]For observations where sales data are available, sales are highly correlated with user ratings.

increases to 1404 products, on average, where the bottom falls out. This is an increase of 846 (s.e. = 104.2) above the 558 (s.e. = 63.1) products, on average, within narrowly-defined subcategories. On the face of it and following conventional intuitions on effects of entry and increases in variety, in terms of numbers of products, we might expect large effects. However, it is important to keep in mind that this is entry by mostly lowest-quality developers into narrowly-defined, already-served product types with established market niche-leaders. (We might expect too that added developers could potentially have an effect on the creation of new niches, but the research design and structure of data examined here does not allow for such an analysis and this remains a question deserving future research.)

If following existing insights on professional complementors working on platforms we might expect the entry of large numbers of complementors might have any number of possible positive or negative effects (externalities), as has been theorized (Economides, 1996; Eisenmann, et al. 2006) and found (Augereau, 2006; Boudreau, 2012) in past research. This question remains to be systematically studied in the case of the bottom-falling-out to amateurs, and is crucial to evaluating any effect on the basket of complementary goods. I divide the following exploration into questions of pricing and possible substitution between amateurs and higher-quality professionals, possible externalities and effects on investment incentives, and observed effects on resulting quality.

5.1 Weak Evidence of Substitution and Competitive Pressure on Top Developers

Panel IV of Figure 7 first seeks to detect whether there is evidence of substitution in pricing patterns. If large numbers of amateurs created crowding and congestion, we might expect this would create downward pressure on the price of higher quality professional developers. As might be expected, Panel IV of Figure 7 shows that top apps tend to be priced highest, with price becoming lower with rank. Broadly speaking, pricing is about the same whether the bottom-falls-out or not for top apps, as the estimated confidence intervals for price versus rank are overlapping. The exception might be among very top apps, where the estimated mean price curve appears to be a tiny bit lower.

To more precisely verify whether there are indeed slightly lower prices, I separately estimate the effect on just a subsample of top-ranked apps across each of the 503 subcategories, controlling for 43 categories. If I compare just the 503 top-ranked apps in each subcategory, the estimated effect is not statistically significant at -0.09 ($s.e. = 0.19$)—or about a dime less, but too noisy to be significant. The effect is weakly statistically significant in estimates on just the top-3 apps or -0.258 ($s.e. = 0.141$). The effect becomes statistically insignificant again among top-10 apps, but the point estimate remains negative at -0.088 ($s.e. = 0.12$)—again, about a dime less, but too noisy to be significant. Therefore, all results point to a weak negative effect on price of the bottom-falling-out to amateurs on highest-ranked developers (cf. Miric, 2017).

5.2 No Evidence of Crowding or Congestion

To assess whether there might exist some form of negative externality, congestion or crowding-out, I seek evidence of any differences in investment incentives, as reflected in versioning. The interpretation here is that—all else being equal—that developers with higher investment incentives will engage in more versioning of their products. Again, as might be expected, levels of versioning are greatest for top-ranked

apps and are lower for lower-ranked apps. The patterns are again very similar for cases where the bottom has fallen out or where it has not, only that the long right tail is much longer where the bottom has fallen out. The estimated relationship between versioning and rank order is statistically identical for the most part, with largely overlapping 95% confidence intervals, as in Panel III of Figure 7. Where there is any possible difference, among very top-ranked apps to the far left of the curve, versioning is in fact slightly higher in the case where the bottom has fallen out to amateurs. (Regression analysis focused on just top apps confirms this point.) Therefore, there is no evidence here that going from hundreds of professional developers to hundreds more amateurs within this context results in any net diminution of investment, as indicated by number of versions of any title. Similar conclusions are drawn if interpreting number of product releases as an indication of willingness to invest or experiment, as in Panel II of Figure 7.

<FIGURE 12>

<FIGURE 10>

5.3 Increase in Higher-Quality Complementary Goods

It remains a question of how adding amateurs impacts the quality of products added to the platform. Patterns studied in Figure 7 would themselves suggest even a possible increase in quality given no diminution in the release of new versions or new products—and even perhaps a slight increase among top developers. Indeed, the earlier Figure 6 provides some suggestion that, while the bulk of developers added were of very lowest quality, there would also appear to be greater numbers of high-quality products. For example, although it is perhaps difficult to immediately discern in Figure 8, 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. Of course, it is possible that lowest-quality and least-reviewed apps have highest variance, leading to many "false positive" when counting seemingly high-quality apps. Table 6 therefore re-estimates the increasing numbers of apps with at least 4.5 out of 5, but sampling only those with some minimum number of user ratings. As reported in the table, whether sampling on >10, >100, >500, or >5000 ratings, I find that the number of highest-quality products roughly doubles the number of highest-quality products across a wide range of estimates. Even estimating the effect of the bottom falling on numbers of top apps with at least 150,000 user ratings (the 99th percentile), I find much more than double (437% times) the greater of high-quality products.

<TABLE 6>

5.4 Possible Interpretations and Areas for Future Research

There are a many plausible mechanisms why the bottom-falling-out could lead to more high-quality, either from products generated by developers who initially entered as amateurs but went on to become professionals, or those originally joining as professionals. Indeed, it is possible that each mechanism operates at the same time. As elaborated below, most deserve closer scrutiny in future research.

1. *"Many Shots-on-Goal & Extreme Value Outcomes."* Large numbers of amateurs with virtually zero chance of a positive income and expected losses might at least in rare instances hit upon a good idea

and become successful. The earlier theoretical analysis would then predict such a developer would then have incentives to invest to a point where it would become a professional, and simply appear in the data here as just another high quality supplier. (A number of studies make analogous arguments in relation to income-seeking professionals, e.g., Terwiesch and Xu, 2009; Boudreau, et al. 2011; Waldfogel 2012, 2014; Waldfogel and Reimers, 2015; Aguiar and Waldfogel, 2016, 2018.) Strictly speaking, a longer panel-based research design allowing some form of "genealogy" of developers is required to investigate this mechanism. Nonetheless, even the analysis provided here at least suggests this explanation is plausible on its face. Consider for example, that as the bottom-falls-out, there are only 2.4 added products with 4.5 ratings with at least 500 reviews (Table 6). On its face, generating fewer than 3 more products by this standard—while adding hundreds upon hundreds of developers as the bottom-falls-out—would still mean a tiny probability of this "extreme value" quality outcome. It should be highlighted however, that perhaps making this seem more plausible still is that even attaining this threshold of quality likely does not on its own imply achieving a commercially successful app. The bulk of share in any given subcategory goes to the top one or two developers. An app with 500 high-quality ratings or more is by no means necessarily commercially success.

2. *"Amateur Learning & Evolution."* Closely-related to arguments above is the possibility that the long-lived amateurs on the platform could at least over time engage in gradual experimentation and learning to eventually achieve higher outcomes than they would upon initial entry. Perhaps a high motivation to learn and insulation from market forces could even create advantages for amateurs to grow and evolve, to deliver these added 2.4 highly-rated products. Again, a longer panel-based research design allowing some form of "genealogy" of developers is warranted and encouraged to investigate this possibility in future study.

3. *"Sticky Information & Heterogeneity."* Reducing the bare minimum costs and "entry barriers" to historically low levels to let the bottom-fall-out creates the opportunity for non-specialists to participate (Jeppesen and Lakhani, 2011; Murray et al, 2016) and perhaps allow a wider range of diversity, and perhaps also even "sticky" knowledge regarding the needs of particular customer groups or novel approaches to addressing problems (e.g., von Hippel, 1998; Jeppesen and Lakhani, 2010; Jeppesen and Frederiksen, 2006). These arguments do not likely explain the particular patterns documented here given that the research design here focuses on narrow precise-designed, establish product categories at a snapshot in time. That is not to say, however, that these arguments were not at work on the AppStore over time.

To the extent a platform can support large numbers of amateur complementors without creating undue crowding or congestion (e.g., Bresnahan et al. 2014), findings across several literatures hint at possible benefits of amateurs on platform outcomes (e.g., von Hippel 1998**,** Jeppesen and Frederiksen, 2006; Jeppesen and Lakhani, 2010; Kittur et al. 2013; West and Bogers, 2014; Waldfogel and Reimers, 2015; Aguiar and Waldfogel, 2016, 2018; Lyytinen, et al. 2016; Sauermann and Franzoni 2015).

1. *"Knowledge Spillovers and Vicarious Learning."* The ideas, examples, failed and successful experiments that are carried out by amateurs might also, in the end, be observed by others—even higher-quality professional entrepreneurs—who then successfully implement these ideas in highest-quality

products. This could be true even if amateurs' own ideas never attain commercial success or even technical success, as ideas may be discerned one feature at a time and adopted. If this were the operating mechanism, then the 2.4 added high quality products could come from the ranks of already relatively high-quality professionals, rather than from the seemingly "utterly long-shot" amateur. In this case, the higher-ranked developers would simply have to observe and learn from other developers. This activity is increasingly becoming easier on online platforms. If this were the case, we might expect an overall improvement and learning in areas where the bottom has fallen out, given there is more productive knowledge to build on for all. Indeed in Figure 6, there is some suggestion of greater numbers of higher quality across the entire distribution of quality, although the differences seem more acute in most extreme high-quality outcomes. The differences therefore appear to manifest most in the extreme upper tail. It is a priori unclear why greatest learning and impact should take place in very best products.

2. *"Lower Experimentation Costs–for all."* At least some of the approaches to platform design that serve to reduce minimum development costs, w_{min}, Table 1 might also plausibly serve to reduce the costs of making higher-cost discretionary investments at the same time, R'_w (e.g., powerful tools). If this were the case, then professionals might simultaneously benefit from lower cost of experimentation and themselves release more products, do more learning and ultimately generate better products—whether amateurs do so or not. This would be additionally plausible if the platform did not incur undue coordination costs, congestion and crowding from amateurs—as appears to be the case, here. The preceding analysis did not test or allow for this possibility. Recall, rather than exploit changes in platform design (here constant within a given platform), the analysis exploited inherent differences across given app types. The econometric analysis found that the proxy for minimum costs, *MinFileSize*, were rather unrelated to higher statistics of file size. (models 3, 4 and 5 of Table 5). Further, the greatest impact on quality appears to concentrate greatest on very top apps and developers – precisely where cost structure of very minimum viable products might be least relevant.

3. *"Strategic Incentives to Invest in Product Development."* In considering the effects of added suppliers—potential competitors—the bulk of existing research, particularly that rooted in industrial economics, tends to consider effects of added entry and competition on strategic incentives (e.g., Aghion, et al., 2005; Anderson and Cabral. 2007). Here, the relevant question is whether the bottom-falling-out to added amateurs could have stimulated incentives to engage in more active development by high-quality developers. This could happen if, on the one hand, there are two-sided network effects that were to create a demand-pull for more innovation an development. On this account, work by Bresnahan, et al. (2015) suggests that the strength of two-sided network effects–at the margin—were quite limited on the Apple AppStore, once the platform grew to a critical scale. On the other hand, it is also possible for the threat of competition, particularly close neck-and-neck competition, to stimulate innovation incentives and racing. Here we see some (weak) evidence of substitution and competitive pressure, in the form of slightly lower prices among top-developers, where the bottom falls out (Section 5.1). For this to serve as an explanation, it would have to be true that these top developers were attempting to "escape competition" from the flood of low

quality amateur developers, which would seem perhaps less likely on its face than other plausible explanations listed earlier.

6 Summary & Conclusion

Crowds of amateurs are often willing to participate in markets, particularly online digital platform-based marketplaces, developing products without any expectation of earning income or revenues. This paper sought to provide a clearer understanding of how platform design influences participation of amateurs and professionals on platforms.

The empirical analysis of mobile apps found, as predicted, that tiny incremental shifts in a proxy measure of minimum costs led the bottom-to-fall-out to amateurs—at least once costs became sufficiently low. Also consistent with predictions, the minimum cost at which the bottom-fell-out was higher (more easily attained) in cases of games, with statistically higher non-pecuniary motivations. Where the bottom fell out, amateurs were not just lower quality, but largely of very lowest observable quality. Nonetheless, they were relatively long-lived on the platform. In these cases, developers tended to generate relatively high numbers of versions and products, relative to their rank-order within their submarkets might otherwise suggest.

The evidence further suggests that the bottom falling out to hundreds of added amateurs creates some weak degree of downward price pressure on top professional developers, but no evidence of net crowding or congestion, in terms of willingness of top developers to generate more products and new versions or higher quality products. In fact, the bottom falling out—despite leading to a flood of lowest-quality products, also increases the number of highest-quality products. I discussed a handful of alternative plausible explanations, most of which cannot be ruled out with the data and research design available here—and which deserve further research.

So, to answer the main question posed by this study, the participation by is determined by a range of dimensions of platform design, which each incrementally move a minimum quality threshold for professionals: minimum development costs (e.g, access fees, etc.), rate at which development costs translate to higher revenues (e.g., tools, etc.), factors influencing non-pecuniary payoffs (e.g., public profiles, etc.), and expected revenues (e.g., contractual restrictions from charging users, etc.). By contrast, amateurs are affected instead only by the interplay of minimum development costs and non-pecuniary motivations, and will join professionals if minimum costs fall lower than non-pecuniary payoffs. To have amateurs on their own, those conditions must hold, while either outright barring professionals access, or otherwise restricting revenues.

The results also clarify that our existing theories and frameworks are useful in studying this new generation of markets and market actors. At the same time, these markets differ from usual textbook descriptions. For example, the "Long-Tail" of suppliers and products on platform-based marketplaces (Brynjolfsson et al., 2003, 2011), will where the bottom falls out be a great deal "longer" than standard theory and intuitions predict—and amateurs are not subject to market discipline in any neoclassical sense. Integration of existing research insights across the largely disparate literatures now studying crowds and platforms—e.g., Computer Science, Economics, Information Systems, Innovation Science, Human-Computer Interface, Law, Management, Network Science, Sociology—is likely to bear considerable fruit.

REFERENCES

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

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. "Competition and innovation: An inverted-U relationship." *The Quarterly Journal of Economics* 120, no. 2 (2005): 701-728.

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

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.

Anderson, Axel, and Luis Cabral. 2007. "Go for broke or play it safe? Dynamic competition with choice of variance." *RAND Journal of Economics* 38.3: 593-609.

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.

Augereau, Angelique, Shane Greenstein, and Marc Rysman. "Coordination versus differentiation in a standards war: 56k modems." *The RAND Journal of Economics* 37.4 (2006): 887-909.

Bakos, Yannis. "The emerging landscape for retail e-commerce." Journal of economic perspectives 15, no. 1 (2001): 69-80.

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.

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

Boudreau, Kevin J. "Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation." *Organization Science* 23.5 (2012): 1409-1427.

Boudreau, Kevin J., Nicola Lacetera, and Karim R. Lakhani. "Incentives and problem uncertainty in innovation contests: An empirical analysis." *Management Science* 57.5 (2011): 843-863.

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

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.

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.

Vancouver

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

Chen, Daniel L., and John J. Horton. "Research note—Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts." *Information Systems Research* 27.2 (2016): 403-423.

Chesbrough, Henry, and Sabine Brunswicker. "A fad or a phenomenon?: The adoption of open innovation practices in large firms." *Research Technology Management* 57, no. 2 (2014): 16-25.

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.

Economides, Nicholas. "Network externalities, complementarities, and invitations to enter." European Journal of Political Economy 12.2 (1996): 211-233.

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. 2016. "The user innovation paradigm: impacts on markets and welfare." *Management Science* 63.5 (2016): 1450-1468.

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, 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.

Glaeser, Edward L., Andrew Hillis, Scott Duke Kominers, and Michael Luca. "Crowdsourcing city government: Using tournaments to improve inspection accuracy." American Economic Review 106, no. 5 (2016): 114-18.

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.

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.

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.

Kohler, T., Fueller, J., Matzler, K., Stieger, D., & Füller, J. (2011). Co-creation in virtual worlds: The design of the user experience. *MIS quarterly*, 773-788.

Krishnan, Vish, and Saurabh Gupta. "Appropriateness and impact of platform-based product development." *Management Science* 47.1 (2001): 52-68.

Lakhani, Karim R. "Managing communities and contests to innovate with crowds." *Revolutionizing Innovation: Users, Communities and Open Innovation* (2017): 109-134.

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

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, How Does Signaling Influence the Rate and Direction of Innovative Activity in a Platform Based Marketplace? (September 5, 2017). Available at SSRN: https://ssrn.com/abstract=3032733 or http://dx.doi.org/10.2139/ssrn.3032733

Miric, Milan, Kevin Boudreau, Lars Jeppesen. 2018. “Protecting their Digital Assets” Unpublished Manuscript

Mollick, Ethan, and Ramana Nanda. "Wisdom or madness? Comparing crowds with expert evaluation in funding the arts." *Management Science* 62.6 (2015): 1533-1553.

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

Muchnik, L., Aral, S. and Taylor, S.J., 2013. Social influence bias: A randomized experiment. *Science*, 341(6146), pp.647-651.

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern. "Of mice and academics: Examining the effect of openness on innovation." *American Economic Journal: Economic Policy* 8, no. 1 (2016): 212-52.

Murray, Fiona, Scott Stern, Georgina Campbell, and Alan MacCormack. "Grand Innovation Prizes: A theoretical, normative, and empirical evaluation." *Research Policy* 41, no. 10 (2012): 1779-1792.

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

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, and Marshall Van Alstyne. "Innovation, openness, and platform control." Management Science (2017).

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.

Powell, Walter W. "A sociologist looks at crowds: Innovation or invention?." *Strategic Organization* 15.2 (2017): 289-297.

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

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.

Ruppert, David, Simon J. Sheather, and Matthew P. Wand. 1995. "An effective bandwidth selector for local least squares regression." *Journal of the American Statistical Association* 90.432: 1257-1270.

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 Zhu. "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.

Su, N., Levina, N. and Ross, J.W., 2016. The long-tail strategy of IT outsourcing. *MIT Sloan Management Review*, 57(2), p.81.

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.

Varian, Hal. 2010. "Computer mediated transactions." *American Economic Review*, 100(2), 1-10.

Verner, June, and Graham Tate. 1992. "A software size model." *IEEE Transactions on Software Engineering* 18.4: 265-278.

von Hippel, Eric. *Democratizing Innovation*. MIT press, 2005.

von Hippel, Eric. "Economics of product development by users: The impact of "sticky" local information." *Management Science* 44.5 (1998): 629-644.

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. von Hippel, Eric. 2017. *Free innovation*. Cambridge, MA: MIT Press.

Waldfogel, Joel. "Copyright research in the digital age: Moving from piracy to the supply of new products." *American Economic Review* 102.3 (2012): 337-42.

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, Joel, and Imke Reimers. 2015. "Storming the gatekeepers: Digital disintermediation in the market for books." *Information Economics and Policy* 3: 47-58.

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.

Zhang, Xiaoquan, and Feng Zhu. 2011. "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia." *American Economic Review* 101.41601-1615.

TABLE 1 Platform Design Choices

Notation	Dimension of Change	Examples
Δw_{min}	Cost structure: Minimum development costs	Reduced platform access charges for complementors; altered development environment (provision of development tools and frameworks, documentation, modifiable examples)
$\Delta R'_w$	Cost structure: Rate that development costs translate to higher revenues	Altered development environment and tools
$\Delta\beta$	Non-pecuniary payoffs	Platform public profiles featuring accomplishments or skills; learning and interaction forums; leader boards
ΔR, Δp	Expected revenues	Complementor licensing programs that proscribe charging users; platform marketing

TABLE 2 Main Variable Definitions

Variable	Unit of Observation	Unit of Measurement	Definition
Endogenous Variables			
NumDevelopers	503 Subcategories	[Developer Count]	The count of unique developers operating and offering products within a given product subcategory
Price	693,541 products	[$]	
Revenues	503 Subcategories	[$000's]	Summation of all estimated annual revenues of all apps appearing in a given subcategory, based on both download charges and in-app purchases
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
NumVersions	693,541 products	[Count]	Number of revisions or new versions of a particular app title that have been released to the sample period
FileSize	693,541 products	[MB]	Amount of memory required to save software app
MeanFileSize	503 Subcategories	[MB]	The mean file size, in megabytes, of all apps appearing within a given product subcategory
Exogenous Variables			
MinFileSize	503 Subcategories	[MB]	The smallest software app, in megabytes, of all apps appearing within a given product subcategory
Games	503 Subcategories	[Indicator]	Indicator switched to 1 for each of the 58 (out of 503) subcategories featuring games
Product Category FE	43 Categories	[Indicator]	Set of fixed effects corresponding to the set of 43 parent categories of the 503 subcategories (e.g., sports games, trivia games, home household and garden, calendars and clocks)
LowMinCost	503 Subcategories	[Indicator]	Indicator switched on for subcategories for which *MinFileSize<.063* (where it is estimated the bottom begins to fall out of the market to amateurs)

TABLE 3 Motivations of Part-Time and Full-Time Developers (Sorted by Difference)

Sources of motivation:	Part-Time	Full-Time	Difference	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. N = 809 developers, in total; 135 part-time developers; values indicate fractions of respondents replying in the affirmative that this motivation is an important.

TABLE 4 Nonlinear Response to Incremental Reductions in Minimum Development Cost

Dep. Var.:	*NumDevelopers*	
	Simple	Piece-Wise Linear Model
Model:	(1)	(2)
MinFileSize		-159
		(358)
MinFileSize × *LowMinCost*		-20,914***
		(2,963)
LowMinCost *(I{MinFileSize<.063})*	494***	1,252***
	(49)	(145)
Constant	362***	381***
	(27)	(60)
Adj-R^2	.09	.16

Notes. Standard errors clustered by app categories are reported in parentheses. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations = 503 app subcategories.

TABLE 5 Robustness Tests

Dep. Var.:	*NumDevelopers*					
	Added Control Variables		Compare Minimum with Other Distributional Statistics			Monte Carlo Re-Sampled *MinFileSize*
Model:	(1)	(2)	(3)	(4)	(5)	(6)
MinFileSize	-706*	-738	-4,291***			-4,421***
	(425)	(985)	(614)			(621)
1stPctlFileSize				736**		
				(289)		
MeanFileSize					-960	
					(940)	
MinFileSize × *LowMinCost*	-21,099***	-21,686***				
	(3,005)	(8,148)				
LowMinCost *(I{MinFileSize<.063})*	1,256***	1,316***				
	(147)	(499)				
Revenues		-.1				
		(.1)				
Product Category FE	Y	Y	Y	Y	Y	Y
Adj-R^2	.21	.21	.12	.07	.05	.13

Notes. Standard errors clustered by app categories are reported in parentheses. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations = 503 app subcategories.

TABLE 6 Regression Results for Numbers of High Quality Products

Depvar:	*Number of Highest Rated Apps, >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. Standard errors clustered by app categories are reported in parentheses. *, ** and *** indicate statistical significance at 10%, 5%, and 1%, respectively. Number of observations = 503 app subcategories.

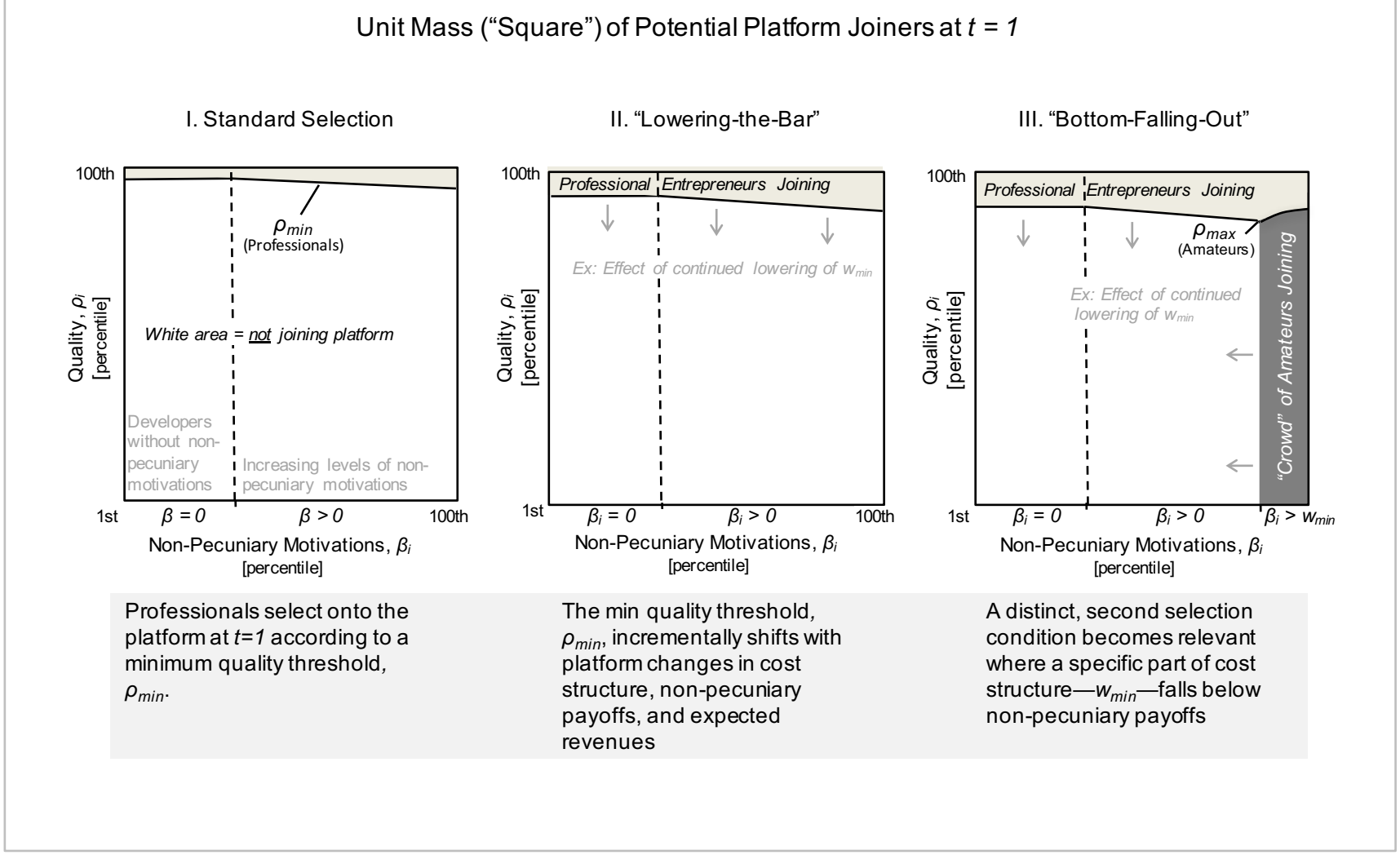


Figure 1 Professionals and Amateurs Join a Platform in Response to Different Thresholds

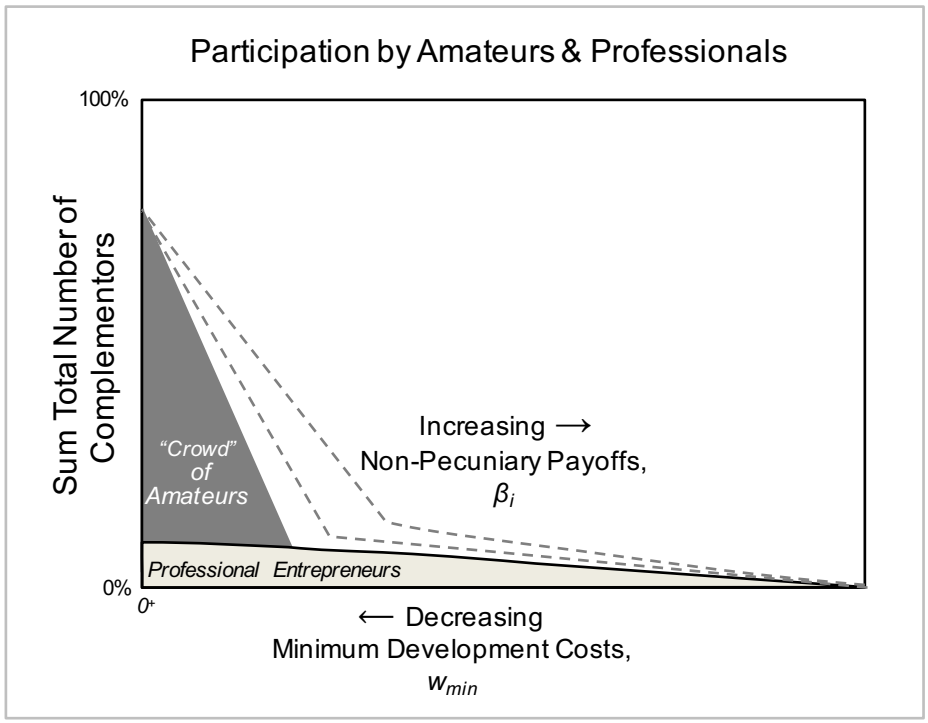


Figure 2 Summary of Main Predictions

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.

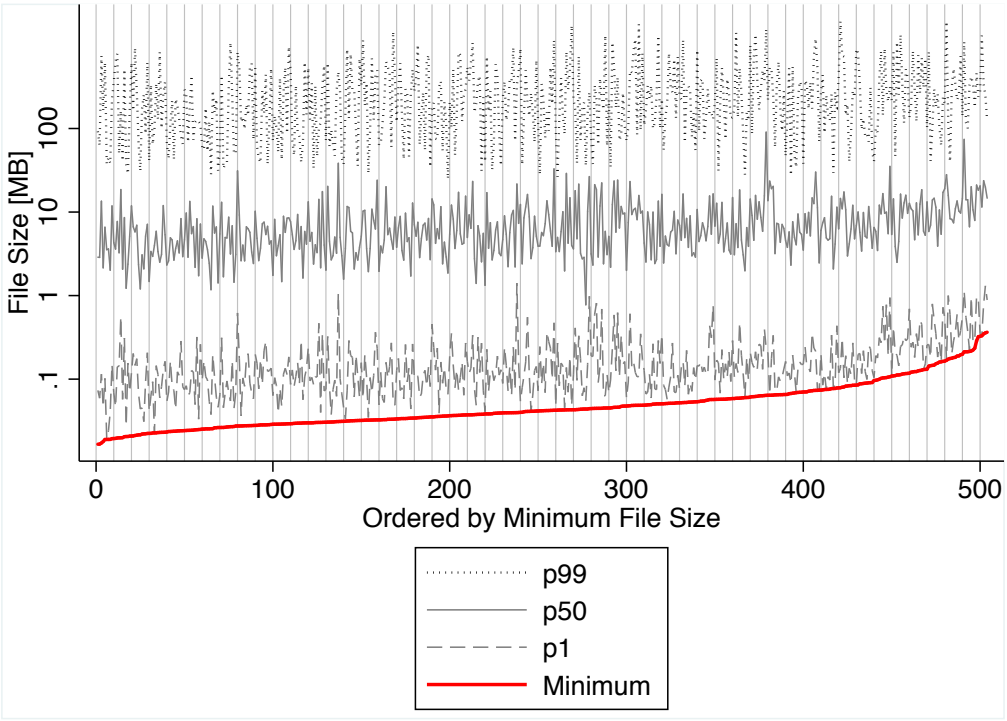


Figure 3 File Size Distribution of Apps Across Subcategories (Ordered by Minima)

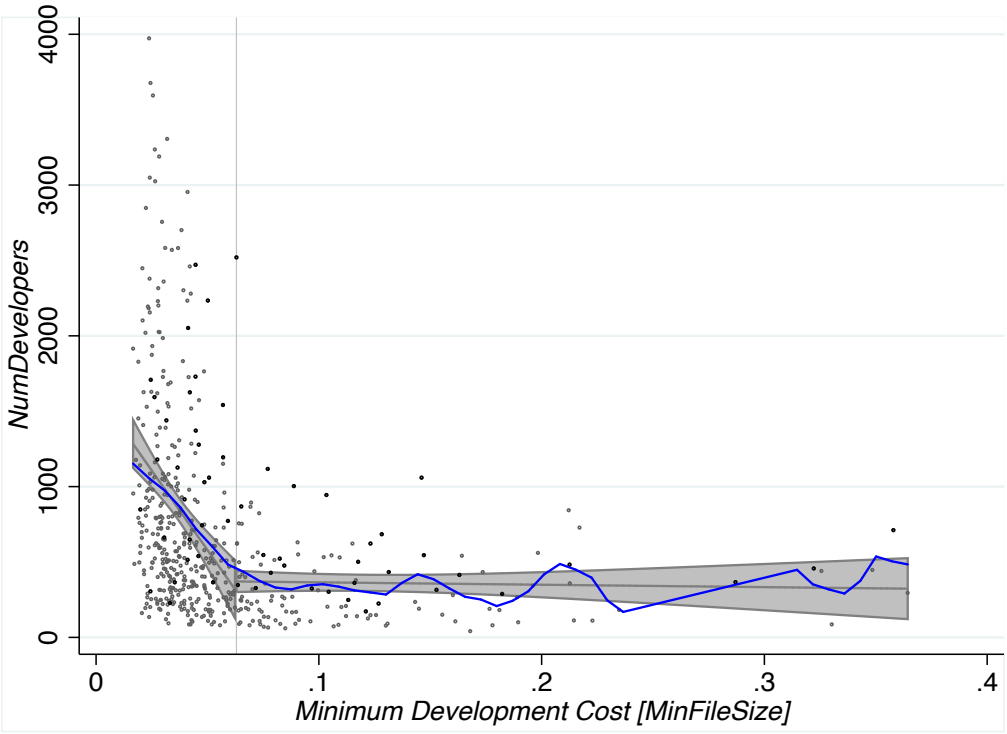


Figure 4 Incremental Changes in Minimum Development Costs and the "Bottom Falling Out"

Note: The flexible nonlinear relationship is estimated with locally-weighted least-squares, weighted by second-order Epanechnikov kernel, with bandwidth chosen according to "direct rule-of-thumb" local-linear method of Ruppert, Sheather and Wand (1995). The piece-wise linar relationship estimates slopes, constants and break-point with maximum likelihood, with 95% confidence intervals shown. Number of observations = 503 app subcategories (i.e., precisely-defined types).

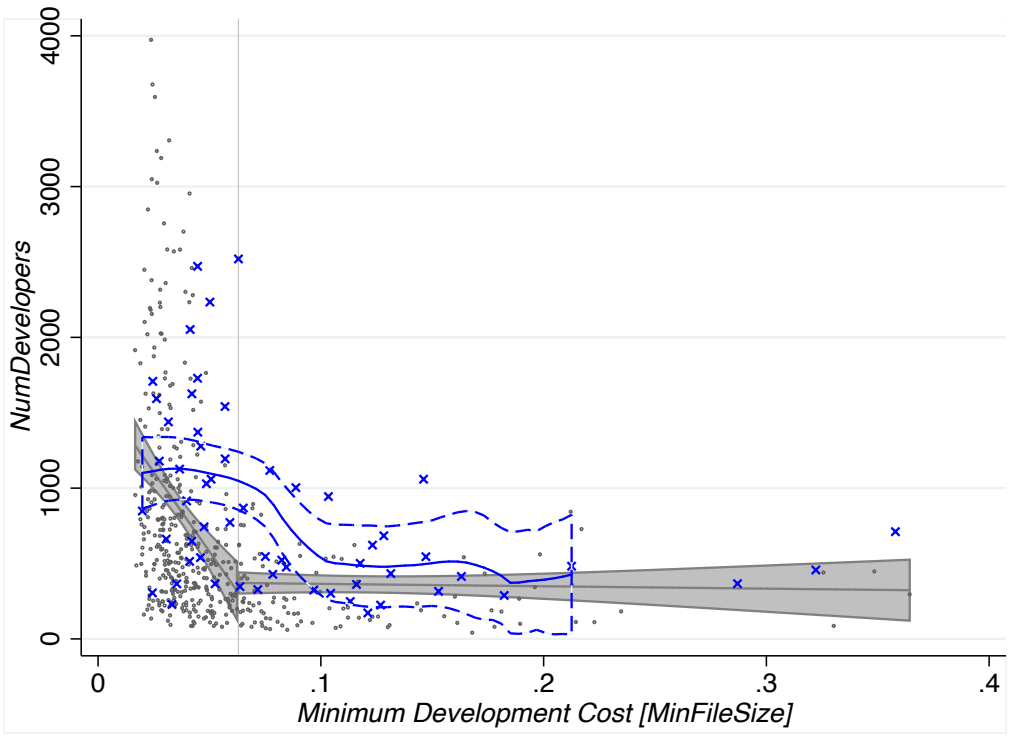


Figure 5 The Bottom Falls Out at Higher Minimum Development Costs for Games (Blue)

Note: The flexible nonlinear relationship is estimated with locally-weighted least-squares, weighted by second-order Epanechnikov kernel, with bandwidth chosen according to "direct rule-of-thumb" local-linear method of Ruppert, Sheather and Wand (1995), with 95% confidence intervals shown. The piece-wise linear relationship estimates slopes, constants and break-point with maximum likelihood. Number of observations = 503 app subcategories (i.e., precisely-defined types), 58 for games.

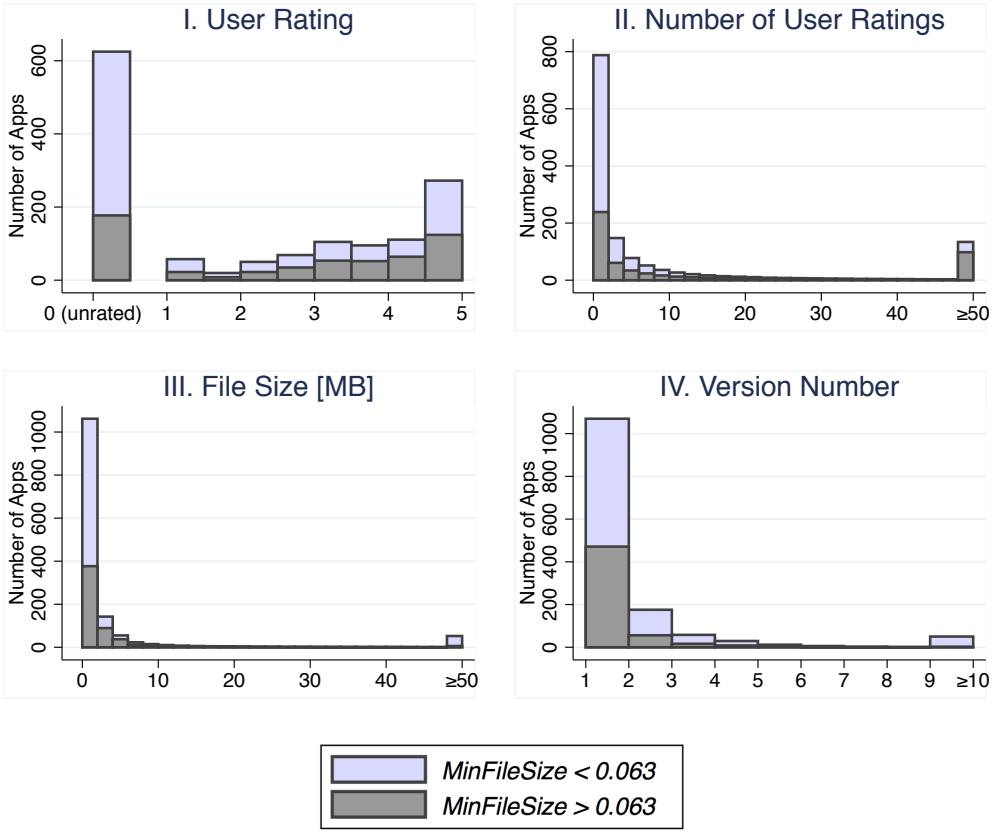


Figure 6 Bulk of Products Added are of Lowest-Quality, as the Bottom Falls Out

Note. Flexible nonlinear relationship is estimated with locally-weighted least-squares. The analysis essentially compares the 503 subcategories at each rank, controlling for 43 categories and contrasting levels across similar subcategories that differ in their *MinFileSize*. The relationships are estimated using the data on all XXXX apps in the population, stratified by subcategories wherein the bottom-has-fallen-out in blue (i.e., *MinFileSize < 0.063*), following the earlier analysis, versus where it has not in grey. The exception are estimates related to numbers of products, which instead of using product-level data use developer-level data (accounting for the different x-axis). Rank order of app demand is approximated using the rank order of number of user ratings. Therefore, for each the top ranks, there are 503 observations for each subcategory. At lower ranks, there are fewer observations as different subcategories have different numbers of developers and products (i.e., the estimates are based on an unbalanced panel, towards the right side).

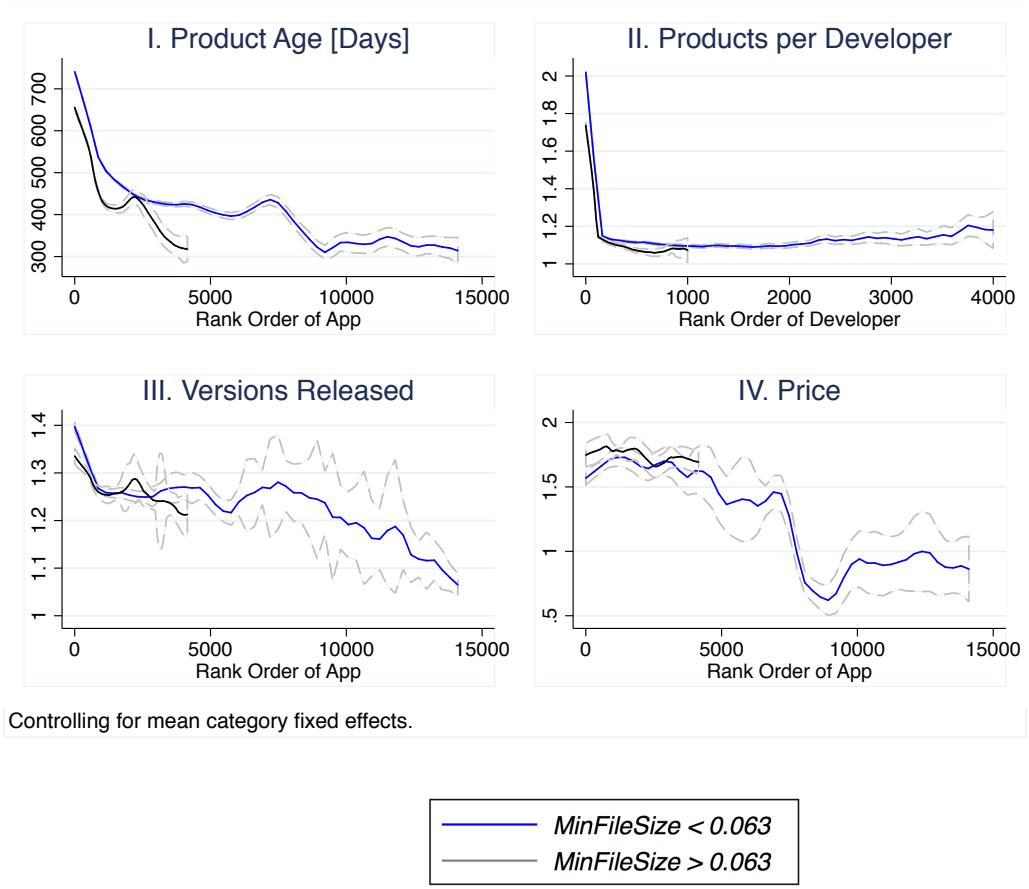


Figure 7 Outcomes across Different Ranks, as the Bottom-Falls-Out (Blue)

Note. Flexible nonlinear relationship is estimated with locally-weighted least-squares. The analysis essentially compares the 503 subcategories at each rank, controlling for 43 categories and contrasting levels across similar subcategories that differ in their *MinFileSize*. The relationships are estimated using the data on all apps in the population, stratified by subcategories wherein the bottom-has-fallen-out in blue (i.e., *MinFileSize* < *0.063*), following the earlier analysis, versus where it has not in grey. The exception are estimates related to numbers of products, which instead of using product-level data use developer-level data (accounting for the different x-axis). Rank order of app demand is approximated using the rank order of number of user ratings. Therefore, for each the top ranks, there are 503 observations for each subcategory. At lower ranks, there are fewer observations as different subcategories have different numbers of developers and products (i.e., the estimates are based on an unbalanced panel, towards the right side). The levels presented are the results, controlling for subcategory means, and then adding back the overall mean levels of each variable to provide a direct estimate of overall mean levels.