KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.1058

"Can a lengthy application title make an application successful?
A perspective of information theory"

Saori Chiba, Chien-Yuan Sher, Min-Hsueh Tsai

April 2021

KYOTO UNIVERSITY

KYOTO, JAPAN

Can a lengthy application title make an application successful? A perspective of information theory

Saori Chiba[a], *Chien-Yuan Sher*[b,*], *Min-Hsueh Tsai*[b]

[a] *Kyoto Sangyo University, Motoyama Kamigamo, Kita-ku, Kyoto, 603-8555, Japan*

[b] *National Sun Yat-sen University, 70 Lienhai Rd., Gushan Dist., Kaohsiung 80424, Taiwan*

[*] Author for correspondence. Address: Department of Business Management, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan. E-mail: cysher@mail.nsysu.edu.tw; Tel: +886-7-5252000 ext. 4629; ORCID: 0000-0002-1444-8450

Declarations

Funding: Kyoto University Foundation and JSPS (KAKENHI 20K01544)

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgements

This paper has benefited from the advice provided by Ming-Jen Lin and the seminar participants at the 2nd Joint Economics Symposium of 4 Leading Universities in Taiwan and Japan in Osaka University, Japan. We are also grateful to the participants at the conference with topic "Business and Management: Framing Compliance and Dynamic" in National Economics University, Vietnam for helpful comments. We appreciate the editorial support from Rahel O'More and Jose de Jesus Herrera Velasquez. Research support from the Kyoto University Foundation and JSPS (KAKENHI 20K01544) are gratefully acknowledged. Sincere gratitude is also extended to the anonymous referees for helpful suggestions. The usual disclaimer applies.

Can a lengthy application title make an application successful? A perspective of information theory

Abstract

We discuss the length limit policies for application (app) titles introduced by the Apple application store (the App Store) and empirically investigate the debate among practitioners with regard to whether lengthy app titles benefit the market performance of the apps. We form communication games between an app seller and a consumer where assembling keywords as app titles provides information for the consumer to find the app matching her needs. A bad type seller stuffs numerous popular but irrelevant keywords to increase the visibility of his app. If the probability that the consumer faces a bad type crosses some threshold, even an informative title composed of relevant keywords cannot transmit more information to the consumer and drive more downloads than a concise title. If this probability is below some threshold, limiting the length of advertisement can hurt the consumer. Based upon random observations of a total of 1,932 apps, we show that before the App Store introduced title length regulations in 2016, even when the title length was longer than 50 characters, this information channel could still be beneficial to the consumer. Therefore, both the 30-character and the 50-character limits may be too stringent, and this restriction would hurt consumers.

Keywords: App Store optimization, keyword stuffing, informational control, average treatment effects
JEL classfication: D83, L86, M37

1. Introduction

In September 2016, the Apple application store (the App Store) introduced a 50-character limit for names of mobile applications (apps) (App Store Review Guidelines History n.d.); this may have caused more than 25% of top apps to change their names because the limit of app names had previously been 255 characters (McCabe 2016). After June 2017, Apple further reduced the limit on app titles to 30 characters. It is believed that this change came in response to 'keyword stuffing' in app titles, a strategy whereby many app publishers were creating lengthy titles with the overall aim of increasing the total downloads of their apps. Keyword stuffing was originally developed in the 1990s as a means of obtaining top search engine rankings through the inclusion of targeted keywords numerous times on a page, and although modern search engines have invalidated such techniques, this idea has been grafted onto the naming of apps.

Keyword stuffing is one popular method of improving the visibility of apps. The App Store is a crowded market; for example, in May 2016, the App Store had a total of 2,309,309 apps available for download, the characteristics of which were quite diverse.[1] It is challenging for app developers to have their apps discovered; it is also difficult for consumers to find a specific app which exactly matches their needs. Practitioners call the process of improving the visibility of apps App Store optimization (ASO). Many ASO marketers believe that creating a proper app title is crucial in ASO (e.g., App Radar n.d.); one common strategy is to bring together numerous keywords as app titles. This practice makes app titles prolix but informative, and can also increase the visibility of apps when consumers use diverse keywords to find apps matching their needs via search engines. Yet, some senior practitioners, such as Patel (2014), argued that lengthy titles make the apps look unprofessional, whilst also creating unpleasant user experiences which can ultimately destroy the willingness of consumers to download the apps. He recommended that the optimal title length should not exceed 25 characters. This raises the interesting question of whether lengthy titles can succeed in driving greater numbers of downloads than more concise titles.

From a perspective of information theory, because it costs consumers to examine the attributes of each app (at least in time if not money), assembling keywords as app titles provides a shortcut for consumers to find the app matching their needs either through search engines or via browsing a list. Therefore, an informative title can drive more downloads than a title with little information. However, if app sellers are free to assemble any keyword as app titles, we can reasonably imagine that some sellers would choose stuffing numerous popular keywords that are irrelevant to their apps to increase the app's visibility. Once the market is full of apps with lengthy and meaningless titles, fooled consumers will finally disregard titles as a channel to identify apps matching their needs; under this scenario, even an informative title composed of relevant keywords cannot transmit more information to consumers and

[1] This count comes from the website, http://www.pocketgamer.biz/metrics/app-store/app-count/ (Accessed June 7, 2016).

hence drive more downloads than a concise title. Moreover, an app store full of apps with lengthy and meaningless titles will also destroy users' aesthetic enjoyment and the titles' usefulness.

We form communication games between a representative app seller and a representative consumer to formalize the discussion above. The seller's app is endowed with multiple attributes. The consumer prefers to download the app if the benefit associated with attributes exceeds the cost. The seller advertises the attribute information (assembling keywords as his app title). He can tell the truth or lie, but his advertisement is constrained to an exogenously determined number of attribute information characters (or a length). The seller is either of good or bad type. The good type is benevolent to the consumer and hence reveals as much information as possible to the consumer. On the contrary, the bad type only wants to induce a download. We find that the bad type makes advertisements noisier, but because of truth-telling by the good type seller, the longest advertisements still transmit more beneficial information to the consumer than shorter advertisements on average and hence lead to more downloads. However, once the probability that a consumer faces a bad type increases, because the channel of information transmission becomes noisier than before, the efficacy of advertising will decrease; in other words, longer titles do not always bring more downloads when the probability of bad types is high.

We also compare models with different lengths of advertisements and find that when the ratio of bad types is below some threshold, limiting the length of advertisements can hurt the consumer because the loss created by limiting the good type seller's information transmission exceeds the gain of preventing noises caused by the bad type seller. Limiting the length of advertisements reduces the channel of information transmission when the ratio of bad type sellers is below the threshold. Increasing the ratio of bad types in the market makes consumers ignore this channel and decreases its efficacy, but the channel sometimes remains partly functional. Interestingly, if we do not consider users' aesthetic enjoyment in the App Store, even though the ratio of bad types is above the threshold, limiting the length of titles cannot improve consumers' welfare. The reason for this lies with the consumer's prior: before the introduction of this character-limit for app titles, consumers have recognized the situation where bad sellers stuff keywords regardless of the characteristics of their apps and have made their own adjustments.

Of course, if a high ratio of bad types causes the channel to become totally dysfunctional, closing the channel of information transmission can save users' aesthetic enjoyment in the App Store.

Our theoretical work is inspired by the current arguments on information design: how different information structures affect different types of agents. For background, see Gentzkow and Kamenica (2014) and Bergemann and Morris (2016). The setup of the limited number of titles (or messages) is relevant to the game theoretic models of optimal organizational languages by Crémer, Garicano, and Prat (2007).

To investigate whether the ratio of bad types is high enough to justify the 30-character limit introduced after June 2017, or the 50-character limit introduced after September 2016, we empirically examine whether the channel of information transmission remained functional when the title length was longer than 30 characters or 50 characters (before the App Store introduced measures punishing apps with long titles and restricting title length in 2016). We randomly sample and observe 1,932 apps in 2015 from an official directory listing all available apps. Our results reveal that apps with a title longer than 50 characters had a better market performance than those with a title whose length was from 30-character to 50-character. We also find that apps with a title longer than 65 characters performed better than apps with title length from 50-65 characters. This suggests that a longer and more informative title could still transmit more information to consumers than a shorter title because the ratio of bad types is not high enough in our sample to disable the communication channel for titles longer than 50 characters. Because the App Store (n.d.) claimed that they always reviewed whether an app title was appropriate for each app, it is reasonable to see that the ratio of bad sellers in the market is not significantly high. Given the relatively small ratio of bad sellers and based on our model, stringent title-length regulations actually hurt consumers.

The remainder of this paper is organized as follows. Section 2 presents our communication games. The materials and empirical method utilized in this study are discussed in Section 3, with Section 4 subsequently reporting the results. Finally, Section 5 concludes.

2. Theoretical Models

We start with a simplified game, the **three-dimensional-advertisement game**, considering a market for travel apps. There are two players, a representative consumer (she) and a representative seller (he).

The seller's app is endowed with a three dimensional-characteristic:

$\theta \equiv (\theta_1, \theta_2, \theta_3)$.

Let Θ denote the set of characteristics where:

$\Theta = \{(0,0,0), (1,0,0), (1,1,0), (1,1,1)\}$.

Each dimension of θ corresponds to a specific attribute, such as types of detailed information that a travel app can provide (public transportation schedule, offline maps, restaurant guide, etc.). For each attribute, 1 means endowment while 0 means no endowment. For example, $\theta = (1,1,1)$ means that the seller's app is endowed with all three attributes. $\theta = (1,1,0)$ means that the app is endowed with the first and second attributes only. The characteristic is drawn from a discrete uniform distribution with support Θ. This distribution is common knowledge. But the realized characteristic is known only to the seller.

The seller launches an advertisement for the consumer; the seller assembles messages or keywords as his app title. The seller's message set is:

$M = \{(\emptyset, \emptyset, \emptyset), (1, \emptyset, \emptyset), (1,1, \emptyset), (1,1,1)\}$.

The seller can select any advertisement m in the message set regardless of his app's characteristic θ. For example, $m = (1, \emptyset, \emptyset)$ can be interpreted to advertise the first attribute only, such as 'Kyoto Public Transportation Guide'. $m = (1,1, \emptyset)$ can be interpreted to advertise the first and second attributes, such as 'Kyoto Public Transportation Guide with Offline Detailed Maps'.

After the seller's advertising, the consumer selects a binary decision, which is given by:

$a \in \{1,0\}$,

where 1 means an action of downloading the app while 0 means no action.

The consumer's payoff is given by:

$a \times (\theta_1 + \theta_2 + \theta_3 - c)$.

The consumer's cost of downloading the seller's app, c, is drawn from a uniform distribution with support $[0, 3]$. The cost is the consumer's private information, and this is realized when the consumer makes a decision.

The first best outcome (for the consumer) is:

$a = 1$ if $c < \theta_1 + \theta_2 + \theta_3$ and $a = 0$ otherwise.

If the first best outcome is not achieved, we say that the consumer's download is insufficient or excessive. Depending on the realized c, there are three possibilities: an app with only the first feature can satisfy the consumer, the consumer needs an app with the first and second features, and the consumer needs an app with all features.

The seller has two types: a bad type ($t = b$) and a good type ($t = g$). The probability of type realization is $p \in (0,1)$ for the bad type and $1 - p$ for the good type.

The payoff of the bad type seller is given by a. That is, the bad type seller prefers the consumer to download his app regardless of the characteristic of his app.

The good type seller is benevolent to the consumer. We assume that the good type seller tells the truth: the good type seller chooses a message:

$$m = \begin{cases} (1,1,1) & \text{if } \theta = (1,1,1) \\ (1,1,\emptyset) & \text{if } \theta = (1,1,0) \\ (1,\emptyset,\emptyset) & \text{if } \theta = (1,0,0) \\ (\emptyset,\emptyset,\emptyset) & \text{if } \theta = (0,0,0) \end{cases}.$$

The good type seller is not strategic (and hence we do not specify the payoff of the good type seller); only the bad type seller and the consumer are strategic.

The timeline is as follows:

1. Nature decides the seller's type (t) and the characteristic of the seller's app (θ). Then, the seller privately and perfectly observes both pieces of information.
2. The seller launches an advertisement (m) for the consumer (i.e., the seller sends a cheap talk

message to the consumer).

3. The consumer decides whether to download the app ($a = 1$) or not ($a = 0$) given her private cost (c).
4. Finally, payoffs are realized for the players.

Our solution concept is perfect Bayesian equilibrium (PBE). The bad type seller's strategy associates the app's characteristic θ with his advertisement m. This strategy is optimal for the bad type given the consumer's strategy and belief. The consumer's strategy associates the seller's advertisement *m* and her cost c with her action a. This strategy is optimal for the consumer given the bad type's strategy and the good type's behavior. The consumer updates her belief using Bayes' rule.

We are going to show **our main claims**:

- First, **the longer advertisement** (i.e., m including more 1s) **results in more downloads** (i.e., *a* = 1 is realized with a higher probability). (Proposition 1)
- Second, **as the probability of the seller having the bad type** (p) **increases, the consumer is worse off.** (Proposition 2)
- Third, **limiting the maximum length of the advertisement never benefits the consumer**. It **hurts her when *p* is small**. (Proposition 3 and Corollary 2)

To describe our claims, we define terminology. Consider a PBE in which there are at least two messages inducing different consumer's beliefs. We call this *an informative PBE*.[2]

Let $\#m$ denote the number of 1s in the message, which we call *the length* (of the message). We say m is longer as $\#m$ is larger.

Proposition 1 *Fix p (the probability of the seller being the bad type) at any level. There is an informative PBE. In any informative PBE, (1) the consumer's belief is equivalent given each m, (2) m including more 1s (the longer advertisement) results in a* = 1 *(a download) with a higher probability, and (3) there is a lower bound L for the length of a noisy advertisement.*

[2] Because we assume truth telling by the good type, there is no off-the-path message.

In any PBE, the consumer chooses $a = 1$ if and only if $\mathrm{c} < \mathrm{E}[\theta_1 + \theta_2 + \theta_3|m]$. Thus, it suffices to describe the bad type seller's message strategy and the consumer's belief and show that both are consistent with each other.

Fix $p \leq 1/3$ (i.e., the seller is less likely to be the bad type). There is a unique PBE where the bad type seller always selects $m = (1, 1, 1)$. The consumer's belief is

$$\mathrm{E}[\theta_1 + \theta_2 + \theta_3|m] = \begin{cases} \dfrac{3+3p}{1+3p} & \text{if } m = (1,1,1) \\ 2 & \text{if } m = (1,1,\emptyset) \\ 1 & \text{if } m = (1,\emptyset,\emptyset) \\ 0 & \text{if } m = (\emptyset,\emptyset,\emptyset) \end{cases}$$

and $(3 + 3p)/(1 + 3p) \in [2, 3)$.

That is, only the longest advertisement $m = (1, 1, 1)$ is noisy in the sense that the consumer does not know whether the seller is the good type telling the truth (with three attributes) or the bad type always exaggerating the features of his app. For any other advertisement, the consumer knows that the seller is the good type telling the truth. On the other hand, the bad type seller always pretends to have three attributes in his app.

Fix $p > 1/3$ (i.e., the seller is highly likely to be the bad type). There are informative PBEs, in any of which the bad type mixes $m = (1, 1, 1)$ and $m = (1,1,\emptyset)$ so that[3]

$$\mathrm{E}[\theta_1 + \theta_2 + \theta_3|m = (1,1,1)] = \mathrm{E}[\theta_1 + \theta_2 + \theta_3|m = (1,1,\emptyset)].$$

The consumer's belief is:

$$\mathrm{E}[\theta_1 + \theta_2 + \theta_3|m] = \begin{cases} \dfrac{5+p}{2+2p} & \text{if } m \in \{(1,1,1),(1,1,\emptyset)\} \\ 1 & \text{if } m = (1,\emptyset,\emptyset) \\ 0 & \text{if } m = (\emptyset,\emptyset,\emptyset) \end{cases}$$

and $(5 + p)/(2 + 2p) \in (3/2,2)$.

Two advertisements, $m = (1, 1, 1)$ and $m = (1,1,\emptyset)$, are noisy. Given these two

[3] There are many strategies of the bad type leading to this belief. For example, regardless of θ, the bad type sends $m = (1, 1, 1)$ with probability x and $m = (1,1,\emptyset)$ with probability $1 - x$, where $x = (1 + 5p)/8p$.

advertisements, the consumer does not know whether the seller is the good type with two or three attributes or the bad type.

Why does the bad type mix advertisements? The bad type selects his advertisement so that it results in the highest belief by the consumer (and hence the most frequent download). Since the good type honestly reveals his information, the consumer trusts every length of advertisement to some extent. Thus, the bad type selects the longest advertisement. But this reduces the consumers' belief in the longest advertisement, and the shorter advertisement may lead to a more frequent download. Hence, the bad type randomly chooses between the two advertisements. This can happen when the seller is highly likely to be the bad type (i.e., p is large).

In each case, the bad type seller's strategy is optimal given the consumer's belief, and the consumer's belief is consistent given the bad type's strategy and the good type's behavior.

Further, the longer advertisement leads to the larger number of downloads on average, and hence the longer advertisement is noisier because the bad type always launches advertisements which induce the largest number of downloads (on average).

The lower bound for the noisy advertisement is $L = 3$ when $p \leq 1/3$ and $L = 2$ when $p > 1/3$. If $\#m \geq L$, the advertisement m transmits noisy information and increasing the length does not result in more downloads. If $\#m < L$, the advertisement m reveals the truth and hence increasing the length results in more downloads. *The long advertisement is too good to be true.*

In other words, the consumer faces noisy information when the seller has the good type with no fewer than L attributes in the app and when the seller is the bad type. Related to this, there are insufficient downloads when the seller is the good type with no fewer than L attributes in the app and excessive downloads (on average) when the seller is the bad type.

From now on, we call the qualitatively equivalent PBEs *the PBE*.

Next, we investigate how the probability of the seller being the bad type affects the outcome.

Proposition 2 *Consider the PBE. As p (the probability of the seller being the bad type) increases, L (the lower bound for the length of noisy advertisement) weakly decreases and the consumer's ex-ante expected payoff decreases.*

According to Proposition 1, larger p results in the same or lower L. Why does this hurt the consumer? First, a larger p makes the noisy advertisement noisier. For example, consider $p < 1/3$. As p increases, the consumer's belief $(3 + 3p)/(1 + 3p)$ given $m = (1,1,1)$ decreases closer to 2. As p increases further, L decreases and more advertisements get noisy.

Now we examine the consumer's welfare. For this purpose, we add terms. $\#\theta$ denotes the number of 1s in the seller's app. Given that the bad type pretends to have no fewer than *j* attributes in his app fixing p, $Pr(\#m \geq j|p)$ denotes the probability that the consumer receives an advertisement where the length $\#m$ is no fewer than *j* (i.e., with length $\#m \geq j$), $E\ [\#\theta|\#m \geq\ j, p]$ denotes the consumer's expected number of attributes when $\#m \geq\ j$, and $Var\ (\#\theta|\#m \geq\ j, p)$ denotes variances of $\#\theta$ when $\#m \geq\ j$.

Fixing p, $V\ (j; p)$ denotes the consumer's ex-ante expected payoff given that the bad type pretends to have no fewer than *j* attributes in his app and $V^*\ (p)$ denotes the consumer's ex-ante expected payoff for the first best outcome. That is, the consumer's ex-ante expected payoff in the PBE is $V\ (L; p)$, $V\ (3; p)$ for $p \leq 1/3$ and $V\ (2; p)$ for $p > 1/3$.

We can show that:

$$V(L; p) = \Pr(\#m \geq L|p)\frac{E[\#\theta|\#m \geq L, p]^2}{6} + (1-p)\sum_{i=0}^{L-1}\frac{i^2}{24}$$

which is equivalent to:

$$V(L; p) = V^*(p) - \frac{1}{6}\Pr(\#m \geq L|p)Var(\#\theta|\#m \geq L, p).$$

Specifically,

$$V(3; p) = \frac{3(1+p)^2}{8(1+3p)} + \frac{5(1-p)}{24}$$

and

$$V(2; p) = \frac{(5+p)^2}{24(1+p)} + \frac{1-p}{24}.$$

Therefore,

$$
\begin{aligned}
&V(2;p) - V(3;p) \\
&\propto \Pr(\#m \geq 3|p)Var(\#\theta|\#m \geq 3,p) - \Pr(\#m \geq 2|p)Var(\#\theta|\#m \geq 2,p) \\
&< 0
\end{aligned}
$$

since $\Pr(\#m \geq 3|p) < \Pr(\#m \geq 2|p)$ and $Var(\#\theta|\#m \geq 3,p) < Var(\#\theta|\#m \geq 2,p)$.

Then, we can show that a larger p reduces the consumer's ex-ante expected payoff. It suffices to consider two cases, case 1 and case 2. In case 1, an increase in p does not affect the lower bound $L = 3$ and:

$$
\lim_{\epsilon \to +0} \frac{V(3,p+\epsilon) - V(3,p)}{\epsilon} = \frac{\partial}{\partial p} V(3,p) = -\frac{1}{2(1+3p)^2} - \frac{1}{12} < 0.
$$

In case 2, an increase in p reduces the lower bound from $L = 3$ to $L = 2$ and:

$$
\begin{aligned}
&\lim_{\epsilon \to +0} \frac{V(2,p+\epsilon) - V(3,p)}{\epsilon} \\
&= \lim_{\epsilon \to +0} \frac{1}{\epsilon} \underbrace{\left(V(2,p+\epsilon) - V(3,p+\epsilon)\right)}_{<0 \text{ (the above mentioned result)}} + \underbrace{\lim_{\epsilon \to +0} \frac{V(3,p+\epsilon) - V(3,p)}{\epsilon}}_{<0 \text{ (from case 1)}} \\
&< 0.
\end{aligned}
$$

Thus, in either case, the consumer is worse off.

The more probable the bad type becomes (higher p), the noisier the advertisement becomes (L decreases) and the consumer is worse off. If p is high, the consumer knows that the seller is highly likely to be the bad type and hence the advertisement is highly likely to be noisy. This knowledge helps avoid excessive downloads when the seller is the bad type. On the other hand, it also makes more advertisements nosier and hence leads to insufficient downloads when the seller is the good type. Overall, the latter negative effect dominates the former positive effect.

Now we introduce **the two-dimensional-advertisement game**. The seller's message set is reduced to:

$$
M_2 = \{(\emptyset,\emptyset,\emptyset), (1,\emptyset,\emptyset), (1,1,\emptyset)\}.
$$

The seller can advertise at most two attributes.

We assume that the good type seller ($t = g$) tells the truth as much as possible in the following sense: the good type seller chooses:

$$m = \begin{cases} (1,1,\emptyset) & \text{if } \theta \in \{(1,1,1),(1,1,0)\} \\ (1,\emptyset,\emptyset) & \text{if } \theta = (1,0,0) \\ (\emptyset,\emptyset,\emptyset) & \text{if } \theta = (0,0,0). \end{cases}$$

Corollary 1 *Propositions 1 and 2 hold in the two-dimensional-advertisement game.*

We can show that in the two-dimensional-advertisement game, for any p, there is a unique PBE that the bad type seller always selects $m = (1,1,\emptyset)$. The consumer's belief in this two-dimensional-advertisement game is:

$$\mathrm{E}[\theta_1 + \theta_2 + \theta_3 | \mathrm{m}] = \begin{cases} \dfrac{5+p}{2+2p} & \text{if } m = (1,1,\emptyset) \\ 1 & \text{if } m = (1,\emptyset,\emptyset) \\ 0 & \text{if } m = (\emptyset,\emptyset,\emptyset) \end{cases}$$

and $(5+p)/(2+2p) \in (2,3)$. In the PBE, the consumer's ex-ante expected payoff is $V(2;p)$. The proof is similar to the one in the three-dimensional-advertisement model.

Only $m = (1,1,\emptyset)$ is noisy. Given this m, the consumer does not know whether the seller is the good type with two or three attributes or the bad type. The good type cannot specify whether he has two or three attributes in his advertisement, and the bad type pretends to have either two or three attributes.

Finally, we argue policy implications of the above mentioned results. Specifically, we are interested in how the policy of limiting apps' lengths affects the consumers.

Proposition 3 *Consider a policy to limit the length of an advertisement from 3 to 2 (i.e. the seller's message set changes from M to M_2). Then, if $p \leq 1/3$, the policy reduces the consumer's ex-ante expected payoff. If $p > 1/3$, the policy does not affect the consumer's ex-ante expected payoff.*

This claim directly results from Proposition 2.

For $p \leq 1/3$ (i.e., the seller is less likely to be the bad type), the lower bound (for the noisy advertisement without policy) is $L = 3(> 2)$, and hence the policy affects the bad type's behavior and makes the consumer's information noisier. By sending the longest advertisement, the bad type pretends to have two or three attributes in his app. The consumer's ex-ante expected payoff is $V(2; p)$. Recall that without the policy, the consumer's ex-ante expected payoff is $V(3; p)$ where $V(2; p) < V(3; p)$. The policy helps to avoid the bad type's exaggeration, but it also prevents the good type's truth revelation and makes more advertisements noisy. The latter negative effects dominate the former positive effects. The consumer is worse-off.

For $p > 1/3$ (i.e., the seller is highly likely to be the bad type), $L = 2$, and hence the policy does not actually affect the outcome (or the consumer's download given t and θ). With or without the policy, the bad type pretends to have two or three attributes in his app and the consumer's ex-ante expected payoff is $V(2; p)$.

Finally, we claim the following.

Corollary 2 *The policy of limiting the length of advertisement is not optimal for the consumer. Furthermore, the policy hurts the consumer more severely when p is smaller.*

In Figure 1, the black solid line, the red dashed line, and the green dashed line represent the total benefit, the change in benefit given good type, and the change in benefit given bad type, respectively, if the maximum length of an advertisement is reduced from three to two.

<Figure 1 about here>

We have proposed a simple theoretical model, but we think these results hold in more general models such as a model with *n* attributes.[4]

---

[4] The analysis of a model with n attributes given a more general discrete distribution is downloadable on https://sites.google.com/a/bu.edu/saori-chiba/home/research

3. Data and Empirical Strategy

According to the model discussed in Section 2, increasing the length of messages does not result in more downloads once the length crosses the lower bound for the noisy advertisement, the level of which depends on the likelihood that the seller is the bad type. Hence, from a consumer's perspective, the 30-character limit introduced after June 2017 is appropriate only if the likelihood that the seller is the bad type is high enough, where titles longer than 30 characters transmit the same amount of information to consumers as titles with 30 characters. In this situation, this 30-character limit does not affect the outcome, and might protect users' aesthetic enjoyment. However, if the ratio of the bad type is low, this policy can hurt consumers. Similarly, whether the 50-character limit is appropriate depends on whether the lower bound for the noisy advertisement is lower than 50-character. Therefore, to verify whether the length limitation is appropriate, we empirically examine whether a longer title could induce more downloads once its length is longer than 30-character (50-character); in other words, we do two empirical analyses: the first one examines whether the lower bound for the noisy advertisement was around 30-character before the App Store introduced title length regulations in 2016, and the second examines whether the lower bound was around 50-character.

3.1 Data

We randomly sample 1,998 apps in 2015 from an official directory listing all available apps.[5] In contrast to previous studies which discussed only the apps in the charts (e.g., Jung et al. 2012, Lee and Raghu 2014), we discuss the features of all available apps in the App Store.[6] Because the app market is very competitive, most apps have never been in the charts.[7] Therefore, a study of the apps in the charts is an

[5] Below is the official list providing the information of available apps in the App Store: https://itunes.apple.com/us/genre/ios/id36?mt=8 (Accessed June 6, 2018).
[6] Google Play Store also imposed similar title length regulations. However, because only Apple provides a public list of all apps, our analysis focuses on the app market of App Store.
[7] Compared with the number of available apps in App Store in May 2016 (2,309,309), the charts only show the top 100, 200 or 300 apps. A survey conducted by a mobile attribution and analytics company, Adjust Inc., pointed out that in the end of 2014, around 83% available apps in App Store appear in no top 300 lists on one-third of their available days (Adjust Inc.

analysis of extremely successful apps only; if the features of successful apps are quite different from those of general apps, then the conclusion of this analysis would be restricted to a small part of available apps in the market.

Since it is costly to observe all the available apps, we make a random sample from the official directory in September 2015, where all the available apps are classified into 23 categories and arranged by their titles in each category.[8] We take a random sample from each category. Table 1 shows the proportion of the number of apps in each category relative to the number of all apps, and the number of apps we pick from each category.

<Table 1 about here>

In this sample, we record both time-variant and time-invariant features for the selected apps; for time-variant features, the observation date is set at December 15, 2015. Our dependent variable, Y_i is an index which indicates whether a selected app was in the chart of *Global Rank* at the observation date.[9] This measure is provided by Adjust Inc.[10] Because Apple does not release data regarding downloads or sales of each app, practitioners usually monitor the App Store rankings as an index of the performance of a particular app in the market; Garg and Telang (2013) showed that researchers could reasonably infer downloads from this ranking data. Furthermore, Carare (2012) demonstrated that whether the selected app was in the App Store chart represented an important sales threshold. Yin et al. (2014) also use whether an app appeared in the top 300 ranking as a measure of success. We therefore regard this index as a reasonable measure of app performance in the market.

Our key explanatory variable is the title length for each app. To examine whether the 30-character limit is appropriate, we divide our sample into three groups: Group 0 contains apps with titles shorter

2015). These kinds of apps, called 'zombie apps,' are effectively invisible to consumers, and can be only found through searching for a specific type of apps or for the app's name.

[8] The category of "Shopping" was created in the end of 2015. In the end of 2016, Apple also created a new category called "Stickers." There were 25 categories in 2017.

[9] The chart of *Global Rank* is created by the mobile attribution and analytics company, Adjust Inc. This chart provides information regarding the global ranking of a selected app if the app is ranked in top 300,000.

[10] This information can be found from the website, http://www.apptrace.com/ (Accessed June 15, 2017).

than 30 characters, Group 1 includes apps whose title length is longer than 30 characters but shorter than 50 characters, and Group 2 includes apps with titles longer than 50 characters. Our main goal is to examine whether an app in Group 2 would have a higher probability of appearing in the chart than that in Group 1. Similarly, to examine whether the 50-character limit is appropriate, we divide our sample into four groups: Group 0 contains apps with titles shorter than 30 characters, Group 1 includes apps whose title length is longer than 30 characters but shorter than 50 characters, Group 2 includes apps whose title length is longer than 50 characters but shorter than 65 characters, and Group 3 includes apps with titles longer than 65 characters. Our main goal here is to examine whether an app in Group 3 would have a higher probability of appearing in the chart than that in Group 2.

Moreover, some of the app titles in our sample are in languages belonging to logographic writing systems which are non-alphabetic, such as Japanese or Chinese. For apps with titles in alphabetic writing systems, the average length of a word in titles is around 6 characters, and hence a title with 30 (or 50) characters can have around 5 (or 8) words. For apps with titles in Chinese, a title with 30 characters can have 15 words (1 Chinese word uses 2 characters). Because the way to count the length of titles in logographic writing systems is quite different from that in alphabetic writing systems, we drop the observations with only a title in logographic writing systems. We also drop a few apps for not having information on some of their features. This results in 1,932 apps in our sample.

Table 2 summarizes the explanations of dependent variables and lists all the covariates we use in the analysis.

<Table 2 about here>

3.2 Empirical strategy

We see our estimation problem in view of the potential outcomes framework (Wooldridge 2010, Cattaneo 2010, Cattaneo et al. 2013). We take the first analysis as an example to explain this empirical strategy. In this analysis, we suppose that for an app *i*, there are three potential values for the outcome: Z_{i0}, Z_{i1}, and Z_{i2}. Z_{i0} is the value of the outcome if *i* has a title shorter than 30 characters. We call this case that i receives no treatment or that i is in treatment level 0. Z_{i1} is the value of the outcome if i has a title whose length is from 30 characters to 50 characters. We call this scenario that i receives level 1 treatment. If i has a title longer than 50 characters, Z_{i2} is the value of the outcome. We call this scenario that i receives level 2 treatment. For app i, the effect of a lengthy title relative to a "common" (or brief) title, which is called a treatment effect, can be defined as the difference between Z_{i1} and Z_{i0} or the difference between Z_{i2} and Z_{i0}.

We observe Y_i, G_i and X_i, where Y_i is the observed dependent variable, G_i denotes the group app i belongs to (or the treatment level i really received), and X_i is a vector containing covariates. We also have three indicators W_{ij}, which take the value 1 if i is in Group j (i.e., $G_i = j$) and the value 0 otherwise. In this framework, the value of Y_i is given by

$$Y_i = W_{i0}Z_{i0} + W_{i1}Z_{i1} + W_{i2}Z_{i2}.$$

Because app i can be in only one group (that is, taking one title), only one of the 3 potential outcomes can be observed. Although two counterfactual potential outcomes for a specific app cannot be observed, under some assumptions, we can still estimate the means of Z_0, Z_1, and Z_2 among apps, or their differences (e.g., the difference between mean of Z_1 and mean of Z_0) (Lechner 2002, Wooldridge 2010, Cattaneo 2010, Cattaneo et al. 2013). The latter is known as average treatment effect. Below are the assumptions we need for the identification:

i. *Conditional independence*: For $j = 0,1,2, Z_j \perp W|X$. In other words, conditional on X, W and (Z_0, Z_1, Z_2) are independent.

ii. *Overlap*: For $j = 0,1,2,$ and all $X \in \chi$, where χ is the support of the covariates, $0 < \Pr(G = j|X) < 1$. In other words, apps of each covariate type always have a strictly positive probability in each group. This means that the treatment i received is not a deterministic function of the covariates.

Based on the assumption of conditional independence, we can get the moment condition below (Wooldridge 2010):

For $j = 0,1,2, \mathrm{E}[Z_j|X, G = j] = \mathrm{E}[Z_j|X]$,

which means that the conditional expectation on Z_j can be identified by conditional expectation of observed outcomes for individuals in that group. Let $\mu_j = \mathrm{E}[Z_j]$ (i.e., the mean of Z_j). Therefore,

$$\mathrm{E}[\mathrm{E}[(Z_j - \mu_j)|X, G = j]] = \mathrm{E}[\mathrm{E}[(Z_j - \mu_j)|X]] = 0. \tag{1}$$

These two assumptions can lead to another moment condition:

$$\text{For } j = 0,1,2, \mathrm{E}\left[\frac{W_j(Z_j-\mu_j)}{\Pr(G=j|X)}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{W_j(Z_j-\mu_j)}{\Pr(G = j|X)}|X\right]\right] = \mathrm{E}\left[\frac{\mathrm{E}[W_j|X]E[(Z_j-\mu_j)|X]}{\Pr(G=j|X)}\right] = 0, \tag{2}$$

where $\mathrm{E}[W_j|X] = \Pr(G = j|X)$. We can also have the third moment condition:

$$\text{For } j = 0,1,2, \mathrm{E}\left[\frac{W_j E[Z_j-\mu_j|X]}{\Pr(G=j|X)}\right] = 0. \tag{3}$$

Through the moment condition (2), Cattaneo (2010) and Cattaneo et al. (2013) proposed the inverse probability weighting (IPW) estimator. In the case of mean of Z_j among apps, the IPW estimator, $\widehat{\mu_{IPW,J}}$, can be obtained from solving the equation below:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_{ij}}{\Pr(\widehat{G_\iota=J}|X_\iota)}(Y_i - \widehat{\mu_{IPW,J}}) = 0,$$

where $\Pr(\widehat{G = J}|X)$ is the estimator for the conditional probability in which app i received level j treatment. Hence, the IPW estimator can be obtained by the weighted mean of the observed outcome through the estimated conditional probability of the received treatment.

From these three moment conditions, Cattaneo et al. (2013) introduced the following condition:

$$\mathrm{E}\left[\frac{W_j(Z_j-\mu_j)}{\Pr(G=j|X)} - \frac{E[(Z_j-\mu_j)|X,G=j]}{\Pr(G=j|X)}(W_j - \Pr(G = j|X))\right] = 0.$$

Motivated by this moment condition, Cattaneo (2010) and Cattaneo et al. (2013) also proposed the efficient influence function (EIF) estimator, $\widehat{\mu_{EIF,J}}$. Below is the analytic solution for $\widehat{\mu_{EIF,J}}$:

$$\widehat{\mu_{EIF,J}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{W_{ij}}{\Pr(\widehat{G_\iota=J}|X_\iota)}Y_i - \left(\frac{W_{ij}}{\Pr(\widehat{G_\iota = J}|X_\iota)} - 1\right)\widehat{Y_{\iota J}}\right],$$

where $\widehat{Y_{\iota J}}$ are the predicted values from regression Y_i on X_i for observations in Group j. In our analysis, we use the user-written STATA command, **poparms**, proposed by Cattaneo et al. (2013) to get the IPW and EIF estimators [Please see Cattaneo (2010) and Cattaneo et al. (2013) for more details].

4. Empirical results

4.1 Analysis 1: Was the lower bound for the noisy advertisement around 30-character?

Table 2 provides the descriptive statistics for all variables. From the right half of the table, we can see that on average, apps in Group 2 had a higher probability of appearing in the chart (top-300k) of *Global Rank* than apps in Group 1. Moreover, apps in Group 1 also had a higher probability than apps in Group 0. It seems that the title length of an app (the number of the messages or keywords the app sends to consumers) was still positively correlated to whether the app can be in the chart even though the length was longer than 30 characters. But, does it turn out to be true after we control for other determinants? We need further analyses.

We first run the user-written STATA command, **bfit**, proposed by Cattaneo et al. (2013) to find the best-fitting model for $\Pr(G = j|X)$. This conditional probability in which an app received level j treatment can be specified by a multinomial logit model. To get the IPW and EIF estimator, we need to find the predicted probabilities for all observations first, while the functional form and estimated coefficients in the multinomial logit model are not our concern. To find the best-fitting model, **bfit** can automatically combine all the covariates we have to form different functional forms, run these candidate multinomial logit models, and select the best model with a minimal Akaike information criteria. Table 3 presents the covariates used for the multinomial logit model chosen by **bfit**.

We also use a similar way to select the covariates used in the estimation of $\mathrm{E}[Z_j|X, G = j]$, which we use to calculate $\widehat{Y_{ij}}$.[11] The covariates used in this logit model chosen by **bfit** are shown in Table 3.

<Table 3 about here>

Table 4 presents our estimation of the mean of Z_j (the expectation of the potential outcome if an app received level j treatment). Because the outcome is whether an app was in the chart of *Global Rank* at the observation date, the mean of Z_j is just the probability of being in the chart if an app receives level j

[11] We note that the covariates used in the prediction of the conditional means in Group 0, 1, or 2 are the same.

treatment. The estimation results shown in Table 4 indicate that the probability of being in the chart is always significantly greater than 0 regardless of the treatment an app receives, and that the probabilities are increasing in the treatment level.

Table 5 reports the estimated average treatment effects, which are the differences between probabilities in two treatment levels. In this table, according to the result from the EIF estimator, we can see that if an app receives level 2 treatment (given a title longer than 50 characters), compared to an app in treatment level 1 (given a title whose length is from 30-character to 50-character), its probability of being in the chart would increase by 0.113. The 95% confidence interval of this estimate does not overlap 0, which suggests that we can reject the null hypothesis that this effect is insignificant. The estimation results from the EIF estimator also indicate that the average treatment effect of changing the treatment level from 0 to 1 is 0.062, and that the average treatment effect of going from treatment level 0 to level 2 is 0.175. These two estimated effects are both statistically different from 0.

Table 5 also reports the estimation results from the IPW estimator. These results also reveal that the average treatment effects are all positively and statistically significant. The treatment effects estimated by the IPW estimator are higher than those estimated by the EIF estimator. Because the EIF estimator is more efficient than the IPW estimator,[12] and it also enjoys the double-robust property (Cattaneo et al. 2013), [13] we employ the results from the EIF estimator.

<Table 4 and 5 about here>

4.2 Analysis 2: Was the lower bound for the noisy advertisement around 50-character?

In Analysis 2, we use the same empirical strategy as that used in Analysis 1, but we divide our sample into four groups. Table 6 provides the descriptive statistics for the dependent variable in each group. From this table, we can see that on average, apps in Group 3 had a higher probability of appearing in the

---

[12] The reason is that the EIF estimator uses both treatment probability and conditional mean models, and the IPW estimator uses only one model.

[13] This property says that to consistently estimate the treatment effects, the EIF estimator only requires either the model for treatment probability ($\Pr(G = j|X)$) or the model for the conditional mean ($E[Z_j|X, G = j]$) to be correctly specified.

chart (top-300k) of *Global Rank* than apps in Group 2. It seems that the title length of an app was still positively correlated to whether the app can be in the chart even though the length was longer than 50 characters.

<Table 6 about here>

Table 7 presents our estimation of the expectation of the potential outcome (that is, the probability of being in the chart) if an app received level *j* treatment. The estimation results indicate that the probability of being in the chart is always increasing in the treatment level. Table 8 reports the estimated differences between probabilities in two treatment levels. In this table, according to the result from the EIF estimator, we can see that if an app receives level 3 treatment, compared to an app in treatment level 2, its probability of being in the chart would increase by 0.188. The 95% confidence interval of this estimate does not overlap 0.

<Table 7 and 8 about here>

5. Conclusions

In this study, we carry out random observations of 1,932 apps available from the App Store in 2015, with our estimation results revealing that for an app in 2015, when its title length was longer than 30 characters, its probability of being in the chart of *Global Rank* would still increase in the length. The same relationship holds even when the length was longer than 50 characters. This empirical result suggests that the lower limit for the noisy advertisement was neither 30-character nor 50-character before the App Store introduced title length regulations in 2016.

Therefore, based on our theoretical model, this implies that the ratio of bad sellers was still below some threshold in 2015, and hence consumers believed that the messages sent by sellers could communicate the features of apps when their titles were longer than 50 characters. We can clearly conclude that the 30-character limit and the 50-character limit may be too stringent, which would hurt consumers.

The App Store Review Guidelines published by the App Store in 2017 claim that,

> *"Customers should know what they're getting when they download or buy your app, so make sure your app description, screenshots, and previews accurately reflect the app's core experience."* (App Store n.d.)

Our analysis implies that the App Review process conducted by the App Store does achieve their goals to some extent. Hence, if the App Review process can efficiently reduce the ratio of bad sellers which stuff irrelevant keywords in the titles of their apps, there is little need for a stringent character-limit for app titles. Based on our theoretical model, when the ratio of bad sellers is below some threshold, the message the sellers send to consumers via their app titles can help consumers make an accurate decision; under this situation, cutting the number of messages or keywords the sellers can send will force consumers to face uncertainty, and then hurt their welfare.

Moreover, it seems that lengthy titles are not a modern phenomenon in human history; during the 17th and 18th centuries, books were produced with very long-winded titles. Genette (1997) argued that book titles have three specific functions: designation, connotations and temptation. The last of these functions may be what leads to such long-winded titles, since longer titles can provide considerable amounts of information regarding the content of a book. The case of book titles in this period and their relevance to the efficacy of app titles may also be worthy of further investigation.

References

Adjust Inc. (2015) The undead app store: the course of discovery in 2015. Report, Adjust Inc., Berlin. Accessed January 20, 2017, https://www.adjust.com/assets/downloads/the-undead-app-store.pdf.

App Radar. App title: Writing Android app titles that drive downloads. Accessed July 15, 2020, https://appradar.com/academy/app-store-optimization-guide/app-title/.

App Store. App Store review guidelines. Accessed September 15, 2017, https://developer.apple.com/app-store/review/guidelines/.

App Store Review Guidelines History. September 01, 2016: Updated subscription rules, Sirikit, Stickers and more. Accessed January 20, 2017, http://www.appstorereviewguidelineshistory.com/.

Bergemann D, Morris S (2016) Information design, Bayesian persuasion, and Bayes correlated equilibrium. *American Economic Review*. 106(5): 586-591. http://dx.doi.org/10.1257/aer.p20161046

Carare O (2012) The impact of bestseller rank on demand: evidence from the app market. *International Economic Review*. 53(3): 717–742. http://dx.doi.org/10.1111/j.1468-2354.2012.00698.x

Cattaneo MD (2010) Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*. 155(2): 138-154. https://doi.org/10.1016/j.jeconom.2009.09.023

Cattaneo MD, Drukker DM, Holland AD (2013) Estimation of multivalued treatment effects under conditional independence. *Stata Journal*. 13(3): 407-450. https://doi.org/10.1177/1536867X1301300301

Crémer J, Garicano L, Prat A (2007) Language and the Theory of the Firm. *The Quarterly Journal of Economics*. 122(1): 373-407. https://doi.org/10.1162/qjec.122.1.373

Garg R, Telang R (2013) Inferring app demand from publicly available data. *MIS quarterly*. 37(4): 1253-1264. http://dx.doi.org/10.2139/ssrn.1924044

Genette G (1997) *Paratexts: Thresholds of interpretation*. Translated by Lewin JE. (Cambridge University Press, Cambridge, England).

Gentzkow M, Kamenica E (2014) Costly persuasion. *American Economic Review*. 104(5): 457–462. http://dx.doi.org/10.1257/aer.104.5.457

Jung E-Y, Baek C, Lee J-D (2012) Product survival analysis for the App Store. *Marketing Letters*. 23(4): 929-941. https://doi.org/10.1007/s11002-012-9207-0

Lechner M (2002) Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society: Series A*. 165(1): 59–82. https://doi.org/10.1111/1467-985X.0asp2

Lee G, Raghu TS (2014) Determinants of mobile apps' success: Evidence from the APP Store market. *Journal of Management Information Systems*. 31(2): 133-170. https://doi.org/10.2753/MIS0742-1222310206

McCabe W (2016) More than 25% of top iOS apps will soon need to change their names. *Sensor Tower* (September 1), https://sensortower.com/blog/new-app-name-length-rules

Patel N (2014) 5 myths about App Store optimization. *KISSmetrics*. Accessed January 20, 2017, https://blog.kissmetrics.com/5-myths-about-aso/.

Yin P-L, Davis JP, Muzyrya Y (2014) Entrepreneurial innovation: Killer apps in the iPhone ecosystem. *American Economic Review*. 104(5): 255-259. https://doi.org/10.1257/aer.104.5.255

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. (MIT Press, Cambridge, MA).

Figure 1 The benefit of consumers if the maximum length of advertisement is ***reduced from three to two***

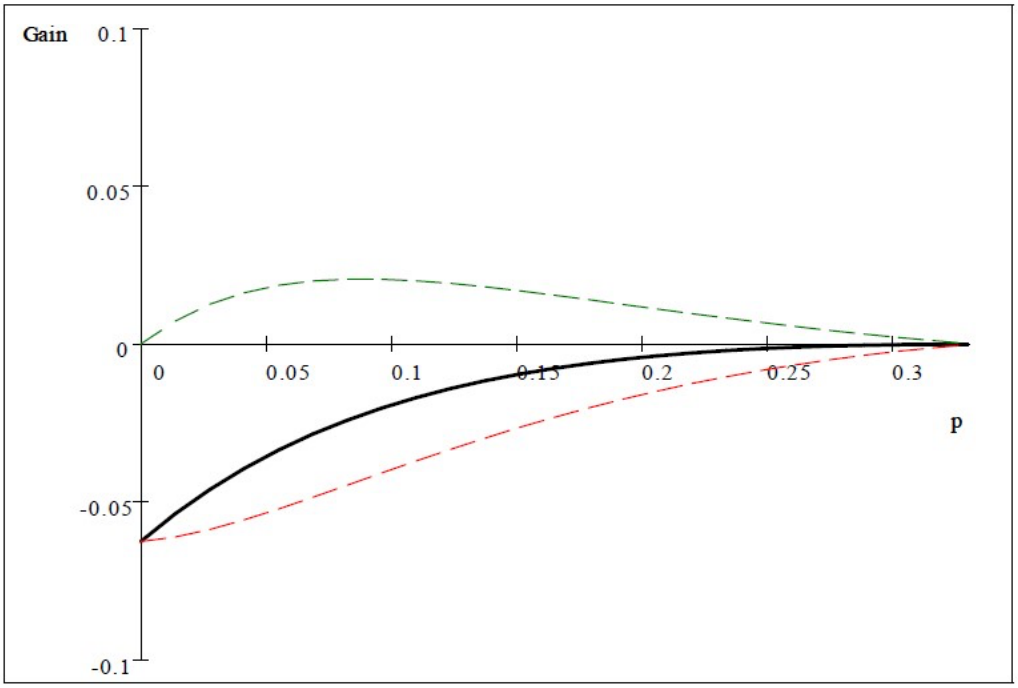

Table 1 The ratio of the number of apps for each category in App Store

Category[a]	Proportion of the category in the population	The number of apps we picked in the category		The number of apps we used in the analysis	
Books	2.93%	59	2.95%	56	2.90%
Business	6.49%	130	6.51%	127	6.57%
Catalogues	0.96%	19	0.95%	18	0.93%
Education	8.51%	170	8.51%	165	8.54%
Entertainment	14.76%	295	14.76%	292	15.11%
Finance	1.99%	40	2.00%	38	1.97%
Food & Drink	2.05%	41	2.05%	37	1.92%
Games	15.00%	300	15.02%	298	15.42%
Health & Fitness	2.87%	57	2.85%	52	2.69%
Lifestyle	9.50%	190	9.51%	175	9.06%
Magazines & Newspaper	0.77%	15	0.75%	13	0.67%
Medical	1.85%	37	1.85%	37	1.92%
Music	2.62%	52	2.60%	47	2.43%
Navigation	1.77%	35	1.75%	34	1.76%
News	2.11%	42	2.10%	41	2.12%
Photo & Video	2.42%	48	2.40%	47	2.43%
Productivity	4.07%	81	4.05%	77	3.99%
Reference	3.51%	70	3.50%	70	3.62%
Social networking	2.27%	45	2.25%	42	2.17%
Sports	2.83%	57	2.85%	56	2.90%
Travel	4.11%	82	4.10%	79	4.09%
Utilities	6.18%	124	6.21%	122	6.31%
Weather	0.44%	9	0.45%	9	0.47%
Total	**100%**	**1998**	**100%**	**1932**	**100%**

Notes:

[a] There are 23 categories in 2015, but 24 and 25 categories in 2016 and 2017, respectively.

Table 2 Lists of variables and descriptive statistics

Group		Full sample		Group 0 [a]		Group 1 [a]		Group 2 [a]	
Number of observations		1932	(100%)	1562	(80.8%)	246	(12.7%)	124	(6.4%)
Variables	Description	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Dependent variable									
Top300k	This app was in the top-300k in the *Global Rank* at the observation date.	0.23	0.42	0.20	0.40	0.31	0.46	0.47	0.50
2. Covariates									
Lite	This app has a lite version.	0.04	0.19	0.04	0.19	0.03	0.18	0.06	0.25
NoUPD	The app has never been updated after the release.	0.52	0.50	0.52	0.50	0.52	0.50	0.52	0.50
Life	The length from the observation date to the day when the app was released (unit: days)	805.61	595.68	818.63	599.89	785.11	589.37	682.35	541.55
Size	The downloading size of the app (unit: megabytes)	32.68	89.21	29.64	86.01	39.51	84.83	57.38	126.14
DEV	Developer scale: the number of available iPhone apps its developer made	52.99	102.88	46.48	95.61	73.39	122.51	94.48	130.93
Free	The price to download this app is zero.	0.72	0.45	0.74	0.44	0.61	0.49	0.69	0.47
IAP	Availability of in-app-purchase option	0.06	0.23	0.05	0.21	0.09	0.28	0.12	0.33
LAN	The number of available languages	2.89	5.60	2.75	5.32	3.29	6.37	3.93	7.12
NoRate	No ratings available at the observation date (information provided by *Adjust Inc.*)	0.63	0.48	0.63	0.48	0.62	0.49	0.55	0.50
Rating [b]	Ratings at the observation date (information provided by *Adjust Inc.*)	3.76	1.04	3.74	1.04	3.79	1.02	3.87	1.02
C-Money	The category of this app is Business or Finance.	0.09	0.28	0.10	0.29	0.05	0.22	0.03	0.18
C-Culture	The category of this app is Book, Food, Music, Photo, Travel, or Navigation.	0.15	0.36	0.16	0.37	0.15	0.35	0.13	0.34
C-Game	The category of this app is Entertainment or Game.	0.31	0.46	0.28	0.45	0.40	0.49	0.41	0.49
C-Health	The category of this app is Health , Medical or Sports.	0.08	0.26	0.08	0.27	0.05	0.22	0.09	0.29
C-Prag	The category of this app is Productivity or Utilities.	0.10	0.30	0.11	0.31	0.08	0.27	0.09	0.29
C-Edu	The category of this app is Education.	0.09	0.28	0.08	0.27	0.11	0.32	0.10	0.31

Notes:

[a] Group 0 includes apps with titles shorter than 30 characters. Group 1 contains apps whose title length ranges from 30 characters to 50 characters. Group 2 includes apps with titles longer than 50 characters.

[b] Descriptive statistics for *Rating* are calculated only for apps with ratings. The value of *Rating* for apps without ratings is set as zero.

Table 3 Covariates used in the treatment model and outcome model

	Model	Covariates and their interaction terms	Akaike information criteria	No. of Obs.
Pr(G=j|X)	Multinomial logit model	Lite, NoUPD, Life, Size, Free, IAP, DEV, DEV^2, DEV*Life, C-Culture, C-Game	2289.33	1932
$E(Z_j \mid X,G=j)$	Logit model	NoUPD, Free, IAP, Size, $Size^2$, LAN, LAN^2, LAN*Size, NoRate, Rating, $Rating^2$, Rating*Size, Rating*LAN, C-Money, C-Culture, C-Game, C-Health, C-Prag, C-Edu	1727.07	1931

Table 4 The estimation of the means of the potential outcomes[a]

	IPW estimator		EIF estimator	
	Coeff.	S.E.[b]	Coeff.	S.E.[b]
If the app received no treatment (Being given a title shorter than 30 characters)	0.202	0.010**	0.205	0.010**
If the app received level 1 treatment (Being given a title whose length is from 30 characters to 50 characters)	0.296	0.027**	0.267	0.027**
If the app received level 2 treatment (Being given a title longer than 50 characters)	0.482	0.038**	0.380	0.038**
No. of Obs.	1931		1931	

Notes:

[a] The outcome is 'whether the app was in the top-300,000 in the *Global Rank* at 12/15/2015'. Therefore, the mean of the potential outcome is the probability that the app was in the chart.

[b] * $p < 0.05$; ** $p < 0.01$.

Table 5 The estimation of average treatment effects[a]

	IPW estimator				EIF estimator			
	Coeff.	S.E.[b]	95% Conf. Interval		Coeff.	S.E.[b]	95% Conf. Interval	
If the app received level 1 treatment vs. If the app received no treatment	0.094	0.028**	0.038	0.150	0.062	0.028*	0.006	0.117
If the app received level 2 treatment vs. If the app received no treatment	0.280	0.039**	0.204	0.356	0.175	0.039**	0.099	0.250
If the app received level 2 treatment vs. If the app received level 1 treatment	**0.186**	**0.046******	**0.096**	**0.277**	**0.113**	**0.046****	**0.022**	**0.204**
No. of Obs.	1937				1937			

Notes:

[a] The outcome is 'whether the app was in the top-300,000 in the *Global Rank* at 12/15/2015', and the mean of the potential outcome is the probability of being in the chart. Therefore, we can see the average treatment effect as the difference between the probabilities of being in the chart in two cases.

[b] * $p < 0.05$; ** $p < 0.01$.

Table 6 Descriptive statistics for dependent variable in Analysis 2

Group	Full sample		Group 0 [a]		Group 1 [a]		Group 2 [a]		Group 3 [a]	
Number of observations	1932	(100%)	1562	(80.8%)	246	(12.7%)	74	(3.8%)	50	(2.6%)
Dependent variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Top300k	0.23	0.42	0.20	0.40	0.31	0.46	0.36	0.48	0.62	0.49

Notes:

[a] Group 0 includes apps with titles shorter than 30 characters. Group 1 contains apps whose title length ranges from 30 characters to 50 characters. Group 2 contains apps whose title length ranges from 50 characters to 65 characters. Group 3 includes apps with titles longer than 65 characters.

Table 7 The estimation of the means of the potential outcomes[a]

	IPW estimator		EIF estimator	
	Coeff.	S.E.[b]	Coeff.	S.E.[b]
If the app received no treatment (Being given a title shorter than 30 characters)	0.202	0.010**	0.205	0.010**
If the app received level 1 treatment (Being given a title whose length is from 30 characters to 50 characters)	0.296	0.027**	0.267	0.027**
If the app received level 2 treatment (Being given a title whose length is from 50 characters to 65 characters)	0.369	0.042**	0.280	0.042**
If the app received level 2 treatment (Being given a title longer than 65 characters)	0.622	0.057**	0.468	0.057**
No. of Obs.	1931		1931	

Notes:

[a] The outcome is 'whether the app was in the top-300,000 in the *Global Rank* at 12/15/2015'. Therefore, the mean of the potential outcome is the probability that the app was in the chart.

[b] * $p < 0.05$; ** $p < 0.01$.

Table 8 The estimation of average treatment effects[a]

	IPW estimator				EIF estimator			
	Coeff.	S.E.[b]	95% Conf. Interval		Coeff.	S.E.[b]	95% Conf. Interval	
If the app received level 1 treatment vs. If the app received no treatment	0.094	0.028**	0.038	0.150	0.061	0.028*	0.006	0.117
If the app received level 2 treatment vs. If the app received no treatment	0.167	0.043**	0.082	0.252	0.075	0.043	-0.010	0.159
If the app received level 3 treatment vs. If the app received no treatment	0.420	0.058**	0.307	0.533	0.263	0.057**	0.150	0.375
If the app received level 2 treatment vs. If the app received level 1 treatment	0.073	0.050	-0.025	0.171	0.013	0.050	-0.085	0.111
If the app received level 3 treatment vs. If the app received level 2 treatment	**0.253**	**0.070****	**0.115**	**0.391**	**0.188**	**0.070****	**0.050**	**0.326**
No. of Obs.	1931				1937			

Notes:

[a] The outcome is 'whether the app was in the top-300,000 in the *Global Rank* at 12/15/2015', and the mean of the potential outcome is the probability of being in the chart. Therefore, we can see the average treatment effect as the difference between the probabilities of being in the chart in two cases.

[b] * $p < 0.05$; ** $p < 0.01$.