Impact of free app promotion on future sales: A case study on Amazon Appstore

Harshal A. Chaudhari
Department of Computer Science, Boston University, Boston, MA 02215, harshal@cs.bu.edu

John W. Byers
Department of Computer Science, Boston University, Boston, MA 02215, byers@cs.bu.edu

Amazon's *Free App of the Day* program, aimed at improving app visibility using daily free promotions, is a compelling experiment in the 'economics of free'. In this study, we investigate the longer-term consequences of free app promotions on the performance of apps on Amazon Appstore. In particular, we quantify the causal impact of such promotions on apps' future download volumes, star ratings, and sales rank using a multi-level model. On average, apps see a surge in download volumes during such promotions, albeit accompanied by a short-term negative effect on its star ratings. On average, sales rank briefly improves but falls to pre-promotion levels within a few months. Interestingly, our findings suggest that lower ranked apps are the biggest beneficiaries of these promotions, as they witness the most significant sales impact. In addition, we show the presence of a cross-market spillover effect of such promotions on the performance of the same apps on Google Playstore. Our results underscore a nuanced set of trade-offs for an app developer: do the benefits of running a promotion and boosting ones' sales rank warrant the lost revenue and risk of lower user ratings in the long run?

Key words: promotion; mobile apps; marketing analytics; Amazon; Google; Android; electronic commerce.

1. Introduction

The introduction of mobile devices such as tablets and smartphones has changed the way in which we work, socialize, and communicate. One important component driving this revolution is the introduction of mobile apps, which allows individuals to do virtually any type of online task from anywhere, from composing e-mail to playing games, to booking a hotel room, to watching live-streamed events. Indeed, an entire ecosystem, the mobile app economy, has been created and has grown to an unprecedented scale in just the past decade. Started in 2007 with Apple's introduction of the iPhone and the Apple App Store, the mobile app economy has grown to a $50 billion economy. In the next five years, the app economy is projected to double in size to over $101 billion, according to market research by App Annie, a website dedicated to the analysis of mobile app stores.[1] This growth has been, and will be driven by the increasing adoption of smartphones around the world.

[1] See: `http://venturebeat.com/2016/02/10/the-app-economy-could-double-to-101b-by-2020-research-firm-says/`

Today, there are four leading app stores: Google Playstore and Apple App Store, each with over two million apps, and more recently, the Windows Store and the Amazon Appstore, an app store for the Android operating system operated by Amazon.com, each with over 600,000 apps[2]. The vast number of apps has two important consequences in such marketplaces. First, search costs are prohibitively high for users to be aware of even a small fraction of apps produced. This incentivies app stores to make their content easily accessible. For example, apps are often divided in topical categories, or are highlighted, *e.g.*, *top charts, early access*, or *editors' choice*. Second, such large volumes generate an intensely competitive environment for app developers, who are often competing for the attention of the same pool of customers. Thus, app developers, in collaborations with app stores or third party companies[3] often advertise their products using both classical and more innovative marketing strategies. These strategies include price-discounted promotions; offering free lite versions of their app; and offering freemium models, where developers provide a game to players free of charge, but then charge a premium fee for special features or content. However, the implications of such promotions are not clear, as each of these options run the risk of losing revenue, to customers who would have paid full price, or would have purchased the premium model, had a discounted version not been on offer. Indeed, it is easy to find blog posts or news discussing the negative effects of such promotions[4].

In our work, we examine one such promotion in detail: Amazon Appstore's *Free App of the Day*, both from the perspective of the Amazon Appstore and the app developers who participated. In this promotion program, on a daily basis, Amazon prominently displays one new paid app from the app store for *free* download in a spot of high visibility on the store website. Clearly, on the day of promotion, a participating app developer suffers short-term losses, as their app is given away for free, presumably including some customers who would have subsequently purchased the app, had the promotion not been in place. But a key selling point of this program that Amazon touts regards *long-term* improvement in sales for apps participating in this promotion. A primary mechanism that could drive future sales is that the promotion causes a significant increase in the short-term popularity of the app, which translates into improved sales rank, which in turn translates into improved placement in Amazon Appstore search results, and better future sales. The extent to which such an effect is operative would be observable within the Appstore itself. A secondary mechanism that could drive future sales is increased awareness and word-of-mouth: the increase in brand and app awareness from a promotion could have a broader secondary effect as new consumers are reached.

[2] See: `http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/`

[3] There are several third party advertising company dedicated to help app developers promoting their apps on different marketplaces. For example, `http://www.appbrain.com/`, creates ad campaigns for the Android market

[4] See:`https://www.developereconomics.com/freemium-apps-killing-game-developers` or `https://gigaom.com/2011/08/02/54805-reasons-not-to-be-amazons-free-app-of-the-day/`

This secondary effect, if operative, would be observable both within the Amazon Appstore, but also in other app stores. A rational app developer, whose goal is long-term revenue maximization, thus has to weigh the short-term downside against the longer-term benefits: for example, assessing whether the incremental revenue from the customers purchasing the app after the promotion and *as a consequence* of the promotion, will offset the revenue lost on the day of promotion, thereby resulting in net profits.

Turning to the perspective of the app store, Amazon Appstore's objectives behind the *Free App of the Day* program are complex, as the Appstore is a two-sided marketplace in a competitive market. From a market structure standpoint, Amazon is the number two player in the Android appstore, market, in direct competition with the Google Playstore, the primary marketplace. But gaining market share against the Google Playstore necessitates becoming more attractive to both sides of the market: in this case, app developers and app purchasers. In some sense, attracting app developers is the easier side of the equation, as it is a relatively low-cost proposition for app developers to multi-home, and market their app in multiple app stores simultaneously. With an increasing customer base and a relatively uncrowded marketplace, Amazon can exploit 'network effects' to attract high quality Android developers to not only publish their apps on Amazon Appstore, but also to use other Amazon cloud services in various functionalities of their app, thereby creating new revenue streams for Amazon. Here too, the Amazon Appstore's incentives are aligned with those of the participating app developers. However, attracting new customers away from Google Playstore may be a more powerful incentive for Amazon to run *Free App of the Day*, as it directly increases market share and also opens up potential revenue streams for Amazon, in the form of app purchases, but also in-app purchases, advertising, and subscriptions. But building share on this side of the market may work against app developers, as doing so prioritizes short-term wins via maximizing free downloads.

Ultimately, the complex set of non-aligned objectives in a two-sided market like Amazon Appstore leaves us with several interesting questions: what are the long-term consequences of participating in deep discount promotions in the Amazon Appstore? Is Amazon's promise of increased post-promotion sales a mere marketing gimmick to convince app developers to participate in the program, or does it hold in practice? What role do various app characteristics play in determining the success of such a promotion? And last, but not the least, does Amazon's promotion strategy have any cross-market effect, on other Android appstores like the Google Playstore? In this paper, we provide preliminary answers to these questions through the lens of a year-long dataset that we collected from the Amazon Appstore and the Google Playstore platforms.

Our analyses show that participation in the Amazon *Free App of the Day* program is positively associated with increased sales volumes on the Amazon Appstore. Higher sales also lead to increased customer reviews. However, they run the risk of attracting customers who review the apps more critically than those who paid full price, much in the spirit of the Groupon effect Byers et al. (2012). We show the presence of a differential impact of promotions on different apps, based on their perceived quality, with low-ranked apps being the biggest beneficiaries of such promotions. We also provide evidence suggesting that extensive marketing campaigns by Amazon does leads to large word-of-mouth and social media engagements for the promoted apps, thereby creating observable spillover effects in other appstores. Our findings extend the understanding of the use of discounted promotions in smartphone app marketplaces. They also complement the existing literature that studies the relationship between performance of a particular app on the appstore and its underlying characteristics. We believe that these results will provide valuable insights to app developers on how to position their apps on the appstores to derive maximum benefit via promotional campaigns.

1.1. Structure of this paper

The remainder of this paper is organized as follows: Section 2 reviews the literature related to our study. Section 3 describes our datasets and theoretical basis underlying our hypotheses. In Section 4, we describe the econometric models used to study our hypotheses. In Section 5, we provide the results, and in Section 6, we describe various robustness checks used to validate our results. Our paper concludes in Section 7 with a review of findings and broader implications.

2. Related Works

Our work connects to several recent streams of research in the marketing community. First, there are a few recent works examining various aspects of the app ecosystem. Notably, Ghose and Han (2014) develop a structural model to estimate consumer demand for mobile apps based by quantifying their preferences for different app characteristics. Shankar and Balasubramanian (2009) provide an extensive review of mobile marketing strategies. Danaher et al. (2011) study mobile phone promotions via coupons, while Bart et al. (2014) conduct a field study to understand the effects of mobile advertisements on consumer attitudes and intentions. Liu et al. (2012) study the impact of freemium strategies on sales volumes and app revenues on Google Playstore, while Cheng and Tang (2010) study similar strategies in software markets. Although, considerable amount of research has been done towards understanding different marketing strategies pertaining to app economy, the abundance of choices, coupled with low cost and minimal learning curve of switching between apps makes it difficult to understand the effects of new marketing strategies like Amazon's *Free App of the Day*. However, the existing literature shows that customers in the app economy make adoption decisions based on two factors: app visibility and app quality.

Anderson (2006) shows that improved visibility is the best way to create demand in a competitive environment. Due to the vast number of apps on popular appstores, the most effective way of improving visibility at no extra cost is by being featured in lists like *highest earning apps, top new* *apps, editors' choices*, etc. All the appstores, including Amazon Appstore, populate many such lists on basis of sales rank, a metric closely related to actual sales volume of an app. Guided by the earliest work of Brynjolfsson et al. (2003) in establishing the relationship between online book sales and sales rank on Amazon.com, researchers have estimated the parameters of the relationship between downloads and sales rank on various appstores using publicly available data (Garg and Telang 2012). These relationships have been used by Chevalier and Goolsbee (2003) to analyze price elasticity and by Ghose and Sundararajan (2006) to study product cannibalization. These studies offer a sound theoretical foundation for hypotheses we investigate in our research.

Another line of related work highlights the economic significance of ratings, rankings, and reviews for both online and traditional marketplaces. Luca (2011) showed that a one-star increase in a restaurant's rating on Yelp results in a 5-9% increase in revenue. Researchers studying Groupon (Byers et al. 2012, Edelman et al. 2016) have shown that while daily deals websites produce a surge of new customers for retail businesses, on average, they negatively impact the reputation of those businesses, as measured through Yelp ratings. Askalidis (2015) studies the impact on sales of large scale promotion on the Apple App Store and Google Play. In contrast with these works, instead of using attributes of customer reviews (i.e., volume and star ratings) to measure the impact of promotion, we additionally use publicly available daily sales ranks. Customers who purchase apps from the appstore may use it for a considerable amount of time, before choosing to review it on the appstore. Using number of reviews as a proxy for app sales may not only introduce a systematic delay in observing the effect of promotion but also mis-estimate its effect, for example by attributing delayed reviews from a promotional purchase to the post-promotion period. For these reasons, we believe that daily sales rank proves to be a more accurate and responsive proxy for estimating sales, as compared to number of reviews.

In addition to product visibility, product quality is an important factor during adoption decision by consumers. On the Amazon Appstore, the visibility and quality of an app are determined by their sales rank, number of reviews, and displayed user ratings. The relatively short life-cycles of apps make it difficult for app developers to build up their brands. Hence, customers usually rely on app characteristics and their ex ante awareness developed via online word-of-mouth, user ratings and reviews, while making purchase decisions. Zhu and Zhang (2010) study the impact of online reviews on the sales of gaming apps, and Chang et al. (1999) study the impact of heterogeneity in customer preferences while making purchase decisions. We employ a similar methodology to these works to ascertain the presence of a similar heterogeneity in the impact of promotion based on the consumer biases in perceived app quality.

Lastly, our study also relates to studies of spillover effects. For example, (Erdem and Sun 2002) empirically studies the cross-category spillover effects of advertising in umbrella brands However, we know of no similar study that empirically observes cross-market spillover effects.

In the next section, we describe our datasets and develop several hypotheses to investigate these factors are affected as a result of promotion.

3. Design of Empirical Study

The *Amazon Appstore for Android* is a third-party appstore for the Android operating system, operated by Amazon.com. It was launched in March, 2011 and is now available in nearly 200 countries. At the time of the launch it had about 3,300 apps; the number has increased significantly since then to nearly 334,000 apps at the time of this study. Similar to Amazon.com, the appstore apps are sold via two channels – website interface and a smartphone app. Amazon.com offers the same selection of apps over both its channels. Because we are unable to distinguish the app downloads over the website channel from the ones over the smartphone app, we are limited to identifying the effects of only the app characteristics that are common to both the channels.

3.1. Free App of the Day Promotion

One of the most high-profile features of the *Amazon Appstore for Android* is the *Free App of the Day*, or FAD. The primary benefit for the apps participating in the FAD promotion is a spot of very high visibility, on both the channels. Along with it, Amazon uses its marketing machinery to promote the participating apps by making Facebook posts or tweets on their official Twitter account. As these posts get picked up by various bloggers and other such platforms, the promotion is only further amplified. The benefits of the promotion continue long after the the app's time in FAD spotlight is over at the end of the day. It finds a spot in the *'Most Recent Free Apps of the Day'* shoveler on both channels.

In addition to the increased direct visibility, the app continues to get post-FAD exposure throughout the appstore due to Amazon's recommendation system. Because of the increase in app downloads typically associated with FAD, the promoted apps show up on the product details pages of other apps under the *'Customers Who Bought This Item Also Bought'* feature. An increase in app downloads also translates into a higher *'Amazon Bestsellers'* list, further improving post-FAD exposure.

3.2. App Selection for Promotion

An interesting feature of the *Free App of the Day* is that the promoted apps are selected by Amazon from the proposals submitted by developers recommending their regularly paid apps for the promotion. Some of the factors taken into account while evaluating proposals are the potential of the app to wide audience, size of the app, number of downloads, plans for marketing outside the appstore, etc. [5] This may introduce considerable selection bias in our analysis which we account for in Section 6.

[5] `https://developer.amazon.com/blogs/post/Tx2CE37E42FQM8M/Submitting-Your-App-for-FAD-Consideration.html`

3.3. Data Description

Our analysis comprises of 3 major datasets.

- Amazon Appstore data.
- FAD promotion history.
- Google Playstore data.

In this section, we provide an overview of the each of the above datasets.

3.3.1. Amazon Appstore Data We collected app profile data from the web interface of Amazon Appstore, while relied upon a third-party Amazon price tracker website `Keepa.com` for collecting daily price and sales rank of every app, from February, 2015 till December, 2015. It includes 23,882 distinct apps from the paid apps sections of the appstore. For every app, we further collected the entire history of publicly displayed user reviews, including submission date, review text, and star-rating, constituting a total of 800,000 user reviews. Thus, our dataset includes *daily* panel data on app sales rank, price, app characteristics and user review data. We capture an exhaustive list of app-related information provided to a user while browsing through the appstore. The observed app characteristics in our sample includes,

- app file size in megabytes,
- app release date,
- app version,
- app description,
- in-app purchase option,
- number of screenshots,
- number of permissions,
- app maturity level,
- app category,
- app developer,
- number of apps provided by the same app developer,
- minimum Android version supported by app.

3.3.2. FAD Promotion History In order to obtain the FAD promotion history, we relied on Amazon Appstore's official Twitter account. Amazon used this account to daily inform its followers about which app was being promoted. There were 794 promoted app over a period of almost two year, from July 2013 to August 2015. To minimize confounding effects of multiple promotions on our analysis, we remove the apps promoted in our observation period, which had already been promoted in past. Thus, for the remaining 179 distinct apps, that participated in FAD promotion exactly once, we record their date of promotion.

Combining Amazon Appstore data with FAD promotion history, we present detailed summary statistics in Table 1 containing detailed summary statistics regarding various app characteristics described above, with 'Treatment' apps corresponding to the ones participating in FAD promotion.

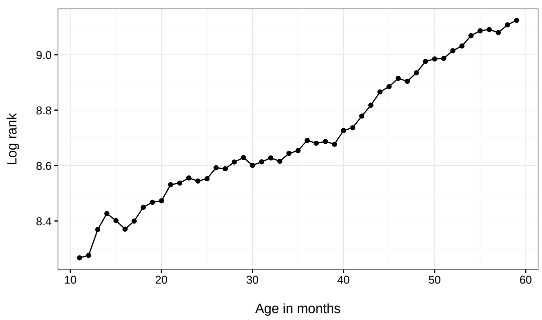
3.3.3. Google Playstore Data We used techniques like image classification on app icon and similarity scores between app and developer names to find cross-listings of FAD promoted apps on the Google Playstore. Out of the original 794 FAD promoted apps, we found cross-listings on the Google Playstore for 720 of them, with a very high confidence. Interestingly, some of the paid from Amazon Appstore are listed as free apps on the Google Playstore. Moreover, due to sheer volume of apps on the Google Playstore, Google does not maintain a sales rank across the entire appstore, only choosing to do so at more meaningful app subcategory level. From publicly available data on `AppAnnie.com` and `AppBrain.com`, we collected sales rank history of 566 of the FAD promoted apps, across 52 different subcategories. For some of these apps, we collected their sales rank history in multiple subcategories. In addition, we also collected a total of 480,000 publicly available user reviews and meta-data for these apps. The summary statistics for Google Playstore data is displayed in Table 2.

3.4. Hypotheses

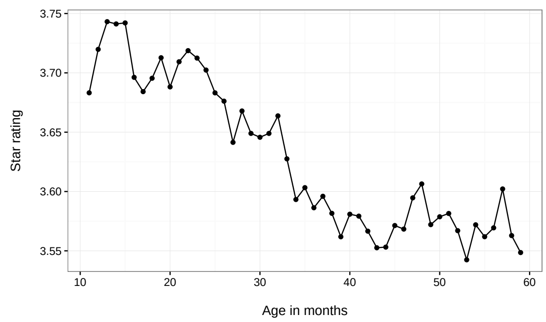
Existing literature in economics and marketing science predicts that the consumers use external information to supplement their ex ante awareness of products while making purchasing decisions (Engel et al. 1995, Kotler and Keller 2006). Amazon's FAD promotion, lowers the cost incurred by the consumers when searching for external information regarding promoted apps, and is thereby expected to affect sales patterns (Brynjolfsson et al. 2006). In this section, we will formulate our hypotheses on how the Amazon's FAD promotion can lead to changes in apps' sales rank and user ratings patterns.

3.4.1. Impact within Amazon Appstore: Amazon's FAD promotion is a unique kind of recommendation tool, that not only provides the product for free, but also decreases the search costs drastically by providing 'directed' links, that take consumers directly to the product pages on the appstore. We hypothesize that such promotions may lead to sales trends with exceptionally high weights for the promoted apps. At the same time, in the spirit of the 'Groupon effect' studied by Byers et al. (2012), we hypothesize that FAD promotion runs the risk of attracting consumers who review the promoted apps more negatively than those who purchase the same apps at full price. Hence, impacts of FAD promotion on longer-term sales and ratings is the central research question studied in this paper.

Now, while describing the various variables from our data, we provide brief theoretical explanation of how they play an important role in the analysis of our hypothesis. The file size of the apps tends



(a) Average sales rank$^{(L)}$ trajectory



(b) Average star rating trajectory

Figure 1 Average sales rank and star rating trajectories on Amazon Appstore

to increase with increasing sophistication and utility, leader to larger download times. Users can incur higher data costs for downloading such apps. This is not only one of the factors influencing Amazon's choice of FAD promoted apps, but may also impact the number of downloads during promotions, even if the users do not have to pay for the app itself. We use the app release date to track the age of an app. As the app gets older, its sales rank tends to increase while average user rating decreases as seen in Figure 1. Furthermore, Amazon prefers to promote relatively newer apps. To maximize the usable variation in different time-variant variables in our dataset, we aggregate them at a monthly frequency, with respect to app age in months. App developers periodically release new versions of their apps to introduce new features or in response to user feedback. Thus, the number of versions of an app is likely to be an indicator of its quality and functional maturity, both of which affect app demand and user ratings. [6]

Prior work by Decker and Trusov (2010) evaluates the relative effect of product attributes and brand names on the perceived product quality in the mobile phone market. Ghose et al. (2009) use image classification techniques on satellite images to infer latent attributes that affect hotel bookings. Similarly, we use the length of the textual description and the number of screenshots provided by the developer on the app profile page to postulate a relationship between quality of app profile and its sales. Android operating system stipulates that apps require explicit permissions from the users to access software and hardware resources of a smartphone like camera, GPS location, memory, image galleries, etc. Chia et al. (2012) find evidence against most apps in misleading users into providing apps with excessive permissions, unnecessary towards the core functionalities of an app. Growing awareness regarding privacy issues potentially affects consumer decisions while installing new apps. Therefore, we include in our analysis, the number of permissions required by an app.

[6] User reviews/ratings on the newer versions of the app may cause an app developer to change certain app characteristics, thereby endogenously determining some of the observed app characteristics. As Amazon does not provide a history of app version updates for every app, we cannot control for these particular unobserved characteristics, although they may be correlated with the observed characteristics.

Amazon rates all apps in its appstore based on information provided by the app developer to determine the appropriate maturity rating. There are four levels of classification – "All Ages", "Guidance Suggested", "Mature" and "Adult". Since, apps of different maturity rating mainly appeal to different segment of users, the maturity rating can have an impact on promotion. Amazon classifies apps into many categories and subcategories for easier search and navigation. We classified the apps in our data into the seven most popular categories – "Games", "Education", "Kids", "Productivity", "Music & Audio", "Photo & Video" and "Other". Because the number of apps vary greatly across categories, we create a variable measuring the number of apps per category and use it as a proxy for the level of competition per category. Some developers are more popular than others on account of the quality and number of apps they publish across the appstore. Large companies developing multiple apps across platforms tend to have much more resources at their disposal compared to indigenous developers and can afford to update their apps more frequently or run their own promotional campaigns. So, we use app developer dummy variables to capture this unobserved heterogeneity between different developers. We also use a variable indicating number of apps published by same app developer to capture the popularity of a developer on the appstore.

Developers of paid apps usually generate most revenue through upfront costs paid by the users at the time of download. However, some apps also include micro-transactions within them allowing users to unlock advanced features and utilities in an app at nominal costs. This 'in app purchase' feature could potentially influence a user's decision while downloading an app even on the day of promotion. Moreover, a developer could also use in app purchase option to lower the upfront cost of the app but recover revenue in future. Minimum Android version supported by app indicates the oldest major Android release version that is completely compatible with the app. Adoption of newer versions has been slow in the Android ecosystem. Hence, supporting older versions of Android could mean targeting larger user base. On the contrary, APIs associated with older versions of Android provide much fewer capabilities. Hence, supporting old versions could drastically limit the quality and sophistication of an app, thereby impacting its downloads and ratings. Because of the multiple confounding factors described above, theory alone cannot predict the impact of FAD promotion on app sales and ratings. Hence, this is an empirical question.

3.4.2. Cross-Market Spillover: In the app economy, consumers are more likely to be aware of popular apps across different appstores, than niche apps specific to their primary appstores i.e., Apple iTunes or Google Playstore. This is because, social media and technology bloggers serve as external sources of information, and play a key role in setting trends for popular apps, as evident in the example of Pokemon Go, a virtual reality based smartphone game (Barnes 2016). We believe that FAD promotion can improve ex ante awareness for promoted apps among the users of Google

Playstore via advertising and word-of-mouth referrals. In fact, Amazon's marketing strategy ensures that consumers can actively perform specific searches on Google Playstore using exact app names. These types of specific searches, or "directed searches" (Moe 2003), take consumers directly to the app profile page, helping them quickly locate it. As a result, we hypothesize the presence of a cross-market spillover effect of FAD promotion on the Google Playstore, on the sales rank of promoted apps. On the other hand, while installing a FAD promoted app, the consumers on the Google Playstore make full priced purchases, and are unlikely to be 'experimenting' like their Amazon counterparts. Hence, we do not expect these consumers to leave overly critical user reviews for their purchases. Consequently, we do not expect presence of a strong cross-market spillover effect on user ratings of the promoted apps.

4. Econometric Model

In this section, we specify the 'within-between' formulation of the multilevel models (Bell and Jones 2015) to estimate the *causal* impact of the FAD promotion on the sales and user ratings' patterns of the promoted apps. In this study, we create a longitudinal dataset by tracking sales rank and review history of many apps over several months. Hence, our study has a hierarchical structure – repeated measurements at level-1, nested individual apps at level-2, which are further nested into separate categories at level-2. While Fixed Effects (FE) models are often considered to be the *gold standard* in such studies, they introduce severe limitations on the effects that can potentially be studied, as they introduce dummy variables corresponding to higher levels of measurement in the hierarchical structure. In this work, estimating the impact of various time-invariant app characteristics on effectiveness of promotions is one of the central research goals. Hence, we use the Multilevel models, which do not have the same limitations. Subsequently, in Section 6, we consider simplified fixed effects models and confirm that the primary effects we identify in the Multilevel case are similarly present (with nearly identical magnitudes) in FE models.

4.1. Model Specification

Because the apps participating in the FAD promotion (treatment apps) were promoted on different days in our observation period, we have a multiplicity of "experiments" to exploit. Our empirical approach relies on contrasting the change in sales rank and user ratings of the treatment apps in a given period with those that did not get promoted in the same period (control apps).

As the age of an app on the appstore increases, its popularity falls due to it being substituted by the newer apps. Hence, the sales rank (user rating) of an app, follow a general trajectory over its lifetime, it increases numerically along with the app age, as evident in Figure 1a. Hence, we adopt an *individual growth model* or level-1 submodel that incorporates this linear change with respect to age of an app. Following the within-between model from Bell and Jones (2015), we also introduce the app-level mean and the centering term for age, a time-varying covariate, to separate the 'within' and 'between' effects of the variable, necessary for causal interpretation. Thus, level-1 submodel is specified as:

$$\text{Sales Rank}_{ij}^{(L)} = \pi_{0i} + \pi_{1i}\text{Age}_{ij} + \epsilon_{ij}{}^{7} \tag{1}$$

where Age_{ij} is a (series of) time-variant value for app i. We stipulate that level-1 residuals are drawn from an underlying normal distribution, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

FAD promotion, lasting for exactly a day, acts as an intervention for the treatment apps, and introduces an abrupt discontinuity in the trajectory of app's sales rank (or user rating) over time. Furthermore, the post-promotion trajectory of the promoted apps may be non-linear over time. Rather than view these patterns as inconveniences, we treat them as opportunities to provide substantiative insights on the effects of the FAD promotion. We begin by describing ways of incorporating abrupt discontinuities into the individual app trajectories, caused by a discrete shock such as a promotion. To postulate such a change, we include a time-varying predictor, After_{ij} in the level-1 submodel that specifies whether and, if so, when each app experiences the discontinuity. Before an app i is promoted, $\text{After}_{ij} = 0$, and if and when, it gets promoted, After_{ij} becomes 1.

$$\text{Sales Rank}_{ij}^{(L)} = \pi_{0i} + \pi_{1i}\text{Age}_{ij} + \pi_{2i}\text{After}_{ij} + \epsilon_{ij} \tag{2}$$

Because After_{ij} distinguishes the pre- and post-promotion epochs for app i, it permits a discontinuity in the *intercept* of the trajectory. The growth parameter π_{2i} captures the magnitude of this instantaneous impact of promotion. To create a post-promotion trajectory, that differs not just in *intercept*, but also in *slope*, we include another predictor Post_{ij} which clocks age of an app from the day of its promotion. Before an app i is promoted, Post_{ij} is 0. On the day the app is promoted, Post_{ij} remains at 0. However, after that, its values begin to increase in concert with the primary temporal predictor, Age_{ij}. It is worth noting that because timing of promotion is app-specific, the cadence of Post_{ij} is also app-specific. Finally, we model the curvilinear change in trajectory post-promotion by adding Post_{ij}^2. The level-1 submodel become,

$$\text{Sales Rank}_{ij}^{(L)} = \pi_{0i} + \pi_{1i}\text{Age}_{ij} + \pi_{2i}\text{After}_{ij} + \pi_{3i}\text{Post}_{ij} + \pi_{4i}\text{Post}_{ij}^2 + \epsilon_{ij} \tag{3}$$

While the level-1 submodel describes how each app changes over observational period, the level-2 submodel we now define describes how those changes differ across apps (Bryk and Raudenbush 1987, Rogosa and Willett 1985). To do so, we introduce app-level means of the time-variant variables

[7] Superscript (L) denotes Logarithm of the variable.

while modeling the *intercept* term. If we let X_i be a vector representing the time-invariant app-specific characteristics, then we can simply include them in the level-2 submodel without the risk of introducing collinearity:

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\overline{\mathrm{Age}}_i + \gamma_{02}\overline{\mathrm{After}}_i + \gamma_{03}\overline{\mathrm{Post}}_i + \gamma_{04}\overline{\mathrm{Post}^2}_i + \alpha\mathrm{X}_i + \zeta_{0i} \tag{4}$$

where $\overline{\mathrm{Age}}_i$ and $\overline{\mathrm{After}}_i$ are the app-level means; as such, the time-invariant component of Age_{ij} and After_{ij} respectively [8]. After combining both the levels of the multi-level model, and some algebraic simplification, we can express a *composite* model as follows,

$$\begin{aligned}\text{Sales Rank}_{ij}^{(L)} = {} & \gamma_0 + \pi_{1i}(\mathrm{Age}_{ij} - \overline{\mathrm{Age}}_i) + \pi_{2i}(\mathrm{After}_{ij} - \overline{\mathrm{After}}_i) \\ & + \pi_{3i}(\mathrm{Post}_{ij} - \overline{\mathrm{Post}}_i) + \pi_{4i}(\mathrm{Post}_{ij}^2 - \overline{\mathrm{Post}^2}_i) \\ & + \pi_5\overline{\mathrm{Age}}_i + \pi_6\overline{\mathrm{After}}_i + \pi_7\overline{\mathrm{Post}}_i + \pi_8\overline{\mathrm{Post}^2}_i \\ & + \alpha\mathrm{X}_i + (\epsilon_{ij} + \zeta_{0i})\end{aligned} \tag{5}$$

where $\pi_5 = \gamma_{01} - \pi_{1i}$, $\pi_6 = \gamma_{02} - \pi_{2i}$, $\pi_7 = \gamma_{03} - \pi_{3i}$ and $\pi_8 = \gamma_{04} - \pi_{4i}$ respectively. Residuals at both levels are assumed to be Normally distributed: $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\zeta_{0i} \sim \mathcal{N}(0, \sigma_0^2)$. Heteroscedasticity at the level-1 is explicitly modeled by including additional level-2 submodel.

$$\pi_{1i} = \gamma_{10} + \zeta_{1i} \tag{6}$$

The residuals part of the *composite* model now becomes $(\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} \times \mathrm{Age}_{ij} - \overline{\mathrm{Age}}_i)$. This reveals two important properties about level-1 residuals: they can be both *autocorrelated* and *heteroscedastic* within-app. Like level-1 residuals, we make an assumption that level-2 residuals have an underlying bivariate normal distribution.

$$\begin{pmatrix} \zeta_{0i} \\ \zeta_{1i} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{pmatrix} \tag{7}$$

The variances associated with level-2 residuals allow us to address how much heterogeneity remains after accounting for the effect of predictors. Because of their conditional nature, the covariance of level-2 residuals, σ_{01}, allows us to address an important question: Controlling for the app-specific time-invariant characteristics, what is the relationship between true sales rank and true rate of growth?

Now, π_{2i} is the 'within' effect and π_6 is the 'between' effect of the FAD promotion (Bartels 2008, Leyland 2010). One of the primary reasons behind preferring a multilevel model over an FE model

[8] This app-level centering is different from centering around the grand mean, which has a different purpose: to keep the value of the intercept of model within the range of the data and to aid convergence. Although, by definition, the app-level mean centering ensures grand mean centering as well.

is its generalizability and extensibility. We outline one such extension by modeling π_{2i} on level-2 to include app-specific time-invariant characteristics. This enables us to model and quantify the effect (if any) of app-specific characteristics on the impact of FAD promotion.

$$\pi_{2i} = \gamma_{20} + \beta \mathrm{X}_i \tag{8}$$

A complex level-1 submodel, such as ours, can end up having many level-2 outcomes. Each level-2 outcome can contain a fixed effect as well as a variance component. Furthermore, there is an added complexity in determining the covariance structure for various variance components in level-2 outcomes. Hence, we sequentially introduce and compare estimated fixed effects and variance components to identify which predictors explain most variation. Similarly, we drop the variance and covariance terms for which null hypothesis cannot be rejected, for example, in Equation 8.

Because the sales rank and user ratings data is observed at a high frequency, serial correlation is a major concern. Following the recommendations of Bertrand et al. (2004), throughout our analysis, we compute standard errors using the generalized Huber-White formula clustered at app level. This allows for arbitrary error correlations among the daily sales rank or user ratings observations. [9]

4.2. Heterogeneous Impact of Promotion

In principle, consumers downloading apps from Amazon Appstore do not make purchase decisions solely based on the price and the app characteristics which are fixed by the app developers; but they may use pre-existing biases or develop some regarding the quality of the apps based on the user reviews and sales ranks of the apps before the day of promotion. For example, we believe that two apps which offer the same core functionality may well experience very different impacts of FAD promotion, if their sales ranks and user ratings are different. Therefore, to check for such a heterogeneity of impact of promotion, we adopt a very conservative definition of "app quality" by segregating the promoted apps into three rank categories based on their average sales rank through our observation period [10]. To each of these rank categories, we add the control apps whose average sales ranks lie within the category boundaries. Table 3 displays sales rank ranges by category.

The model to study the heterogeneity of impact of FAD promotion is as follows,

$$\begin{aligned}\text{Sales Rank}^{(L)}_{ij} = {} & \gamma_0 + \gamma_1(\text{Age}_{ij} - \overline{\text{Age}}_i) + \gamma_2(\text{After}_{ij} - \overline{\text{After}}_i) \\ & + \gamma_3(\text{Post}_{ij} - \overline{\text{Post}}_i) + \gamma_4(\text{Post}^2_{ij} - \overline{\text{Post}^2}_i)\end{aligned}$$

[9] The explicit modeling of heteroscedasticity in level-1 accounts for only a specific kind of heteroscedasticity, when it is linear in nature. The Huber-White formula provides conservative estimates of standard errors without making any assumptions on the nature of heteroscedasticity. Since, true nature of heteroscedasticity is always unknown, we explicitly model the linear part, and allow the rest to be controlled using Huber-White standard error computation.

[10] As shown in the robustness checks section, our findings are robust to the choice of number of categories

$$
\begin{aligned}
&+\gamma_5\overline{\text{Age}}_i + \gamma_6\overline{\text{After}}_i + \gamma_7\overline{\text{Post}}_i + \gamma_8\overline{\text{Post}^2}_i + \gamma_9\text{Rank Category}_i \\
&+\gamma_{10}\text{Rank Category}_i \times (\text{Age}_{ij} - \overline{\text{Age}}_i) \\
&+\gamma_{11}\text{Rank Category}_i \times (\text{After}_{ij} - \overline{\text{After}}_i) \\
&+\gamma_{12}\text{Rank Category}_i \times (\text{Post}_{ij} - \overline{\text{Post}}_i) \\
&+\gamma_{13}\text{Rank Category}_i \times (\text{Post}^2_{ij} - \overline{\text{Post}^2}_i) \\
&+(\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} \times \text{Age}_{ij} - \overline{\text{Age}}_i) \qquad (9)
\end{aligned}
$$

The co-efficients for each rank category in γ_{11} represent heterogeneity of the immediate impact of FAD promotion, while ones in γ_{12} and γ_{13} represent the longer term impact due to the interaction of RankCategory_i with Post_{ij} and Post^2_{ij} variables.

4.3. Cross-Market Spillover

Because we do not have access to the data for control group of apps on the Google Playstore, we cannot specify a model that provides causal inference regarding cross-market spillover effect of FAD promotion. Hence, our model aims at identifying the correlational effects of FAD promotion and sales rank trends on the Google Playstore. Unlike the impact of FAD promotion on Amazon Appstore, we do not expect to longer duration impact on the Google Playstore. Hence, to maximize the usable variation, we code various variables at weekly frequency. Because each app is promoted on a different day on the Amazon Appstore, we create a categorical variable, Interval_{ij} that measures the offset in weeks from the day of promotion, for app i. This allows us to model the spillover effect as follows,

$$
\begin{aligned}
\text{Sales Rank}^{(L)}_{ij} &= \beta_0 + \beta_1\text{Interval}_{ij} + \beta_2\text{AppId}_i + \beta_3\text{AppCategory}_i + \beta_4\text{Time}_j \\
&+\beta_5\text{AppId}_i \times \text{Time}_j + \epsilon_{ij} \qquad (10)
\end{aligned}
$$

where we have included fixed effects for each app, each week, as well as an interaction between the two. Similar to the previous model, we compute standard errors using generalized Huber-White formula, clustered at app level.

5. Empirical Results

Our empirical analysis focuses on four main questions:

- What is the impact of FAD promotion on the sales rank, number of reviews and user ratings of the promoted apps on Amazon Appstore?
- Do the time-invariant app characteristics of the promoted apps have an effect on the potential impact of FAD promotion?
- Is there a heterogeneity in the impact of FAD promotion based on 'quality' of the promoted app?

- Is there a cross-market spillover effect of FAD promotion on Google Playstore?

In this section, we provide a formal analysis of these questions using the models specified in the previous section. Our overall approach starts by fitting the model from Equation 5, with estimates reported in Table 4. Column (1) represents the impact of promotion on sales rank, while columns (2) and (3) represent the impacts on number of monthly reviews and user ratings of promoted apps respectively.

5.1. General trends across apps in the Amazon Appstore

Before describing the effects of FAD promotion specifically, we first quantify general trends across the broad set of apps within the Amazon Appstore. Consistent with our beliefs from Section 3.4.1, we see that as an app gets older, its popularity decreases. Every passing month, the sales rank of an average app falls (numerically increases) by close to 3%[11]. On average, an app in our panel also receives 5% fewer reviews every passing month. Moreover, the star rating associated with a review itself decreases by an average of 0.01 star every month – depreciation of sorts. Interpreting the co-efficient of the 'between' effect associated with Age, we conclude that there is a significant variation between these trends of sales rank, number of monthly reviews and star ratings among different apps. Interestingly, correlation between Age and the *intercept* term of the regression is negative, which implies that, in our observation period, the older apps typically start out with better sales rank, but a lower number of monthly reviews and star ratings.

How app characteristics impact general trends In Section 4, we describe how the use of multilevel models enable us to study the impact of various time-invariant app characteristics on the performance of the app. Point estimates from Table 4 confirm our hypotheses that app characteristics indeed impact the general trends described in the previous section. Notably, we find that increasing file size of sophisticated apps, better textual and graphic description of the app on app profile page, lower maturity rating and higher frequency of updates positively impact the sales rank, monthly reviews as well as the user rating. For a detailed description on the individual effects of these characteristics, please refer to Appendix 8.1. Some of our findings contradict the findings of Ghose and Han (2014) on Google Playstore. It remains to be investigated if the effects of app characteristics vary across different appstores.

5.2. Impact of FAD promotion

Now, we study the impact of FAD promotion on the apps participating in the program. We describe the short term as well as the long term impact on sales rank, monthly reviews and the user rating.

[11] As the popularity of an app declines, its sales rank worsens, corresponding to a numerical increase in rank. Hence, when the dependent variable is Sales Rank$^{(L)}$, a positive co-efficient associated with any predictor implies that app's sales rank falls, while a negative co-efficient implies that sales rank improves with a unit change in the predictor

The co-efficient of the After variable quantifies the immediate impact, experienced during the day of promotion, while co-efficients of Post and Post^2 variables help us better understand the longer term impact by describing the shape of the post-promotion trajectory of the dependent variable. We find that FAD promotion causes a 25% improvement immediately. However, post promotion, the sales rank starts falling at a significantly faster rate than it would have in the absence of any promotion. Comparing the co-efficients of Post and Post^2 variables, we observe that it takes around 3-4 months for the rate of fall of sales rank to stabilize to its pre-promotion rate.

To estimate whether the improvement in sales for a small period after the promotion is enough to offset the losses sustained due to free give-aways of the app during promotion, it is important to know the exact parameters of the Pareto distribution relationship between sales rank and actual sales volume. Estimating these parameters is beyond the scope of this study due to unavailability of actual sales data. Our analysis provides a framework to easily evaluate the net developer revenue, given these parameters. However, it should be noted that a net negative developer revenue does not always mean that a developer would suffer losses. Without conducting a counter-factual experiment, it is likely that we would overestimate the revenue lost at the time of promotion as it is impossible to know how many customers who downloaded the app for free would have otherwise purchased the app at its full price and contributed to the lost revenue.

Consistent with the sales rank trend, we observe that FAD promotion causes an abrupt 18-fold increase in the number of monthly reviews. Similar to sales rank, the number of monthly reviews keeps decreasing until after 4-5 months, at which point they stabilize to the pre-promotion values. However, consistent with our hypothesis, we find that the increased downloads in the month of promotion and subsequent abrupt increase in user reviews is achieved at a cost of a significant decrease in the average star rating. FAD promotion causes an abrupt decrease of 0.16 stars immediately after promotion, and increases the overall rate of decline of star ratings by up to 0.01 stars more every month.

We offer two potential explanations for the decline of star ratings: this could be because the users who download apps during FAD promotion are more likely to be experimenting with new apps. Such users may install an app simply because it is free, notwithstanding their actual needs, and review the app with low rating due to the app's perceived inability to impress them. An alternative explanation is offered via anecdotal evidence[12]. In case of apps that provide services via cloud infrastructure, the overwhelming increase in app usage during promotion may lead to poor quality of service due to inadequate resources, resulting into dissatisfied users who leave critical reviews with low star ratings.

[12] https://blog.shiftyjelly.com/2011/08/02/amazon-app-store-rotten-to-the-core/

Observing the co-efficients of the interaction terms, we find that some of the app-specific time-invariant characteristics affect the effectiveness of the FAD promotion. A 10% increase in app size results in 1.42% fall in the post-promotion immediate sales rank and a 1% fall in the number of monthly reviews in the month of promotion. One extra screenshot in the app profile page improves the sales rank immediately after FAD promotion by 4.5%. Similarly, a 10% increase the length of textual description also improve the effectiveness of FAD promotion by up to 2%. While the price of the app or presence of in-app purchase options within the app significantly affect the general trends on Amazon Appstore, interestingly, they do not affect the effectiveness of FAD promotion.

While our results provide insights into the impact of FAD promotion, they do not provide a conclusive answer to the question faced by the app developers - is it beneficial to participate in the FAD promotion? Our analysis reveals that FAD promotion positively impacts sales ranks and the volume of reviews of the promoted apps, at a cost of significant decline in star rating, underlining a nuanced set of trade-offs for the app developers. However, they indicate that app developers contemplating participation in the FAD program should provide enough textual and graphic information on app profile page and limit the size of app to maximize the benefits of the promotion.

5.3. Heterogeneous impact of FAD promotion

To study the heterogeneous impact of FAD promotion, we fit the model described in Equation 9. Table 5 provides the estimates of fitting this model. As before, column (1) represents the impact on the sales rank, while the columns (2) and (3) represent the impacts on number of monthly reviews and user ratings of promoted apps respectively. We observe a significant heterogeneity in the impact of FAD promotion across different rank categories defined in Section 4.2.

Promoted apps belonging to category 3 benefit the most in terms of immediate improvement in the sales rank after FAD promotion. While, on an average apps in categories 1 and 2 achieve 17.5% and 25% improvement respectively, those belonging to rank category 3 experience an astonishing 54% improvement in their sales rank immediately after promotion. This trend is evident in the number of monthly reviews as well. Compared to category 1, apps belonging to category 2 and category 3 receive roughly 2.5 times as many reviews in the month of promotion. However, this increased user engagement comes at the cost of decreased user ratings. Apps in category 1 suffer a loss of 0.08 stars on average, those in category 2 lose 0.2 stars, while those in category 3 lose 0.35 stars in the month of FAD promotion. We also observe some heterogeneity in the longer term impacts of FAD promotion within different rank categories. Notably, the sales rank of promoted apps belonging to category 2 fall at a significantly faster rate than others. This trend is also mirrored in terms of post-promotion monthly reviews. The number of monthly reviews for promoted apps from category 2 decrease at a significantly faster rate than other categories.

From the perspective of an app developer, it is a surprising finding that low ranked apps are the biggest beneficiaries of promotion in terms of downloads. We speculate that the popular top apps have already cannibalized the market and thus have less potential new customers to attract than relatively unknown low ranked apps, which could explain the observed trends. At the same time, low ranked apps are inherently bad, and their exposure to wider audience via promotion leads to overwhelming number of critical reviews, and the subsequent steep decline in star ratings.

We also provide a visual summary of the heterogeneity in the impact of FAD promotion across rank categories in figures 2 through 4. These figures plot the means of estimated dependent variables for each of the rank categories at 30-day intervals in 5 months prior through half a year following the FAD promotion. The dashed lines in each figure represent robust 95% confidence intervals (allowing for arbitrary within-app error correlations) for each point estimate. These figures provide further evidence supporting our hypothesis that FAD promotion indeed impacts apps of different 'quality', differently. It is evident from the figures that apps belonging to the Bottom Third i.e., rank category 3, benefit the most in terms of improvement in sales rank and increase in monthly reviews, while they also suffer the biggest drop in the star ratings, after promotion. It is also clear from the widths of post-promotion confidence intervals that point estimates are rather noisy. Moreover, these models do not include the full set of covariates used in estimating the Equation 5.

Analysis of the heterogeneous nature of the impact of FAD promotion allows us to look at the question of participation in FAD promotion in a more parsimonious manner. It is evident that top app developers experience an increase in the number of downloads with close to no decline in star ratings, while the developers of the bottom ranked apps, notwithstanding the decline in star ratings, benefit a lot in short-term profits, as a result of increased sales after promotion. We believe that for developers belonging to either of these categories, the improvement in sales through FAD promotion is worth the risk of long-term damage to the app reputation. However, the developers of the apps in middle category face a difficult choice to prioritize either short-term profits or long-term app reputation, with no definite solution.

5.4. Cross-Market Spillover

We now consider the impacts of spillovers from FAD promotion on the Amazon Appstore to other appstores, namely Google. As described in Section 3.3.3, due to a large volume of apps, Google does not publicly display a uniform sales rank for every app across the entire appstore, choosing to do so only at the level of categories. Unfortunately, this limits our ability to quantify the magnitude of the cross-market spillover, as we cannot normalize the effect of FAD promotion across different subcategories without detailed information regarding app downloads for all apps. Hence, we only provide a visual summary of the cross-market spillover effect of FAD promotion on Google Playstore in figure 5, obtained by fitting the model from Equation 10. We plot the estimates (β_1) for the categorical variable Interval which represents the offset in weeks from the day of promotion of an app on the Amazon Appstore along with a 95% confidence interval. We see evidence of an improvement in the categorical sales rank in the week of promotion, supporting the hypothesis of a cross-market spillover effect.[13] The effects seems to last for a few weeks after the FAD promotion. One should not make strong inferences from this figure, however, for the reasons described above. Moreover, due to absence of control apps in the dataset, we cannot demonstrate a causal relationship between the FAD promotion and the observed effect. Nevertheless, we believe that the presence of such a striking trajectory of sales rank for different apps that are promoted on Amazon Appstore, exactly in the week of promotion, is likely a strong indicator of cross-market spillover and warrants further examination. This cross-market spillover effect is also supported anecdotally, e.g., the statistics provided by the developer *Tasharen Entertainment* in their blogpost detailing their experience during FAD promotion[14]. In Figure 6, we observe that, on an average, the star ratings of the FAD promoted apps on the Google Playstore are lower after the promotion, similar to the decline seen on the Amazon Appstore. However, we observe that the star ratings drop occurs roughly two weeks *before* the FAD promotion on average, an effect for which we do not have a ready explanation.

Although, we do not provide *causal* evidence supporting the cross-market spillover effect, our analysis and anecdotal evidence strongly supports the presence of such an effect. A plausible explanation of this spillover effect is that Amazon's aggressive marketing of the promoted apps is an attempt to attract new users to Amazon Appstore. However, after the end of FAD promotion, users who become aware of the promoted app perform 'directed searches' of the app names on their primary appstore i.e., Google Playstore to download the app, instead of downloading Amazon Appstore app and then purchasing the app over it. It remains to be investigated whether spillover effects on Google Playstore mirroring those on Amazon (i.e., improved sales rank but lower ratings) play a role in the developers' decision to participate in Amazon's FAD promotion.

6. Robustness Checks

We implemented a series of robustness checks and found our results are robust to these modifications.

6.1. Multilevel models vs. Fixed Effects models

The technical problems associated with analyzing hierarchical data are well known. Standard OLS regression models, with residuals are independently and identically distributed assume that any two

[13] In February, 2017, Google announced that starting then, it would take into account social media engagement along with sales during calculation of sales rank in a blogpost. However, our study was conducted years before this announcement. https://android-developers.googleblog.com/2017/02/welcome-to-google-developer-day-at-game.html

[14] `http://www.tasharen.com/?p=4664`

higher-level entities (apps) are identical, and can be 'completely pooled' into a single population. However, with hierarchical structure, especially with temporal dependent variables, this assumption is patently false.

The multilevel models solution involves partitioning residual variance into two components: higher-level variance between higher-level entities (apps) and lower-level variance within this entities, between occasions, achieved by having a separate residual term at each level. As such, this is 'partially pooling' the data, as it assumes that all higher-level entities, while not identical, come from a single distribution, which can be estimated from the data itself. However, the standard formulation of multilevel models lead to *exogeneity* assumption; that the residuals at each level are independent of the covariates. However, this assumption often does not hold true in standard multilevel models. The *endogeneity* often arises because effect of every time-varying covariate has two components: called 'between' effect (specific to higher-level entities that does not vary between occasions) and 'within' effect (specific to difference between occasions, within higher-level entities). Standard multilevel models make an unreasonable assumption that 'within' and 'between' effects are equal in magnitude. When these effects are in fact different, some variance is left unaccounted for, thereby leading to the residuals being correlated with the time-varying covariate.

The fixed effects solution to this problem is simple; control out all the higher-level variance, and with it any 'between' effects, by including dummy variables for every higher-level entity i.e., 'no pooling'. Hence fixed effects models only estimate the 'within' effect, at the cost of being unable to estimate effects of higher-level time-invariant covariates (app-specific time-invariant app characteristics like maturity rating, app category, etc.) because all the degrees of freedom at the higher-level are consumed with the dummy variables. In a study such as ours, where time-invariant variables are of particular interest, this is a major limitation.

However, the 'within-between' formulation of the multilevel models (Bell and Jones 2015) provide an elegant solution to this problem by explicitly modeling out the 'between' effect. They achieve this by simply including higher-level mean terms in the standard multilevel models. This removes the correlation between time-varying covariate centered around higher-level means (app) and the higher-level (app) variance. Furthermore, the app-level mean is not constrained by occasion-level effects, thereby letting it completely account for the entire higher-level variance[15]. Thus, multilevel models not only solve the problem associated with *endogeneity* in a more elegant manner than fixed effects models, but are also far more generalizable and extendeable by allowing for complex covariance structures that can model occasion-level heteroscedasticity explicitly or introducing cross-level interaction terms[16].

[15] Demidenko (2013) interprets fixed effects model as simply multilevel models where higher-level variance is infinite.

[16] While fixed effects models can also estimate interactions, they can only be interpreted properly in presence of higher-level means.

In Table 6, we provide a comparison between multilevel model and fixed effects model fitted over our dataset. Columns (1) and (2) of the table provide estimates for the multilevel model and the fixed effects model respectively. For brevity, we have dropped estimates of individual fixed effects corresponding to every app. It should be noted that the multilevel model, in its simplest form, provides numerically exact estimates as the fixed effects model for the impact of FAD promotion. In addition to the estimates provided by the fixed effects models, the multilevel models are able to measure the impact of time-invariant app-specific characteristics as well as the 'between' effects of various time-varying covariates.

In our analysis, we use a more generalized version of the multilevel model, controlling for occasion-level heteroscedasticity and serial correlation using more complex covariance structure of the residuals, leading to more robust estimates of the impact of FAD promotion.

6.2. Controlling for Amazon Selection Bias via Sample Matching

The analyses in Section 5.3 demonstrates that there is causal impact of FAD promotion on the performance of promoted apps' sales ranks, reviews and user ratings. Similarly, there is a heterogeneity in the impact of promotion on account of perceived biases regarding the 'quality' of the app. However, one could argue that since Amazon decides when and whether to promote an app, they systematically only promote apps which have high likelihood of experiencing improvement in sales rank. This Amazon selection bias could have confounded the results from Section 5. We use the propensity score matching method suggested by Rosenbaum and Rubin (1983) to control for this Amazon selection bias.

In our study, we use the propensity score matching method to match Amazon FAD promoted apps (treatment apps) with the control apps, on basis of their various observable characteristics. We focused on matching the treatment and control apps along characteristics that provided least overall standardized bias across all the covariates[17]. A visual comparison between the matched and unmatched sample is provided in Figure 7. We also provide the summary statistics of the matched samples of control apps in Table 7. On comparing with the overall summary statistics of Amazon Appstore from Table 1, we find that the matched set of control apps is no longer significantly different from the promoted (treatment) apps.

We re-estimate the models for impact of FAD promotion (Equation 5) and the heterogeneity of impact (Equation 9) using the matched sample of control apps. The results are presented in Table 8 and Table 9 respectively. We find that results remain qualitatively same regardless of the use of matched sample of control apps. The estimated impact of FAD promotion is about the same

[17] We use the PSMATCH2 propensity score matching module in Stata to match each treatment app with its 20 nearest neighbors. The results are robust to using different numbers of neighbors.

even after controlling for Amazon's selection bias. Due to low propensity of promotion of apps with higher maturity rating, we find that the impact of maturity rating on the general trends in Amazon Appstore becomes insignificant when the matched sample of control apps is used. Furthermore, the 'between' effect associated with age of the app also becomes insignificant as the disparity between the treatment and control apps is low in the matched sample. Even after using the matched sample, we still observe heterogeneous impact of FAD promotion, with a slightly decreased magnitude in case of sales rank and number of monthly reviews, and a 5% decrease in significance for the user rating.

Hence, we conclude that the observable impact of FAD promotion, and its heterogeneous nature persists, even after controlling for Amazon's selection bias using sample matching, thereby providing more evidence supporting our initial hypotheses.

7. Conclusions

Appstores, like most traditional and online market platforms, are dominated by a few best-selling apps, while the large amount of other apps compete for visibility and attention of customers. However, in case of appstores, the absence of operational costs associated with inventory management, and the relative ease of running large scale online advertising campaigns has given rise to very innovative marketing strategies that involve short-term full price discounts. In this paper, we have examined a number of hypotheses to analyze the impacts of deep discounted promotions in the app economy. While there remain challenges in trying to exactly quantify the expected profit/loss margins of such promotions, we provide a framework to do so, conditional to the parameters of the Pareto distribution relationship between sales rank and actual sales volume on Amazon Appstore.

Our empirical results, presented in Tables 4 and 5, highlight that on average, all apps promoted in the Amazon *Free App of the Day* program experience a significant immediate improvement in the sales on account of improved visibility. However, the long-term effects of such a promotion strategy depend on the quality of the app. The improvement in the post-promotion sales volumes may not be sustained long enough to offset the lost revenue on the day of promotion, especially for the top apps. App developers should be cognizant that promotions lead to an abrupt increase in engagement of the users in form of reviews (both positive and negative), and on an average, cause negative impact on the reputation of app. Overall, our study yields very important insights into the successful implementation of deep discounted promotions in the mobile app market. It suggests that, long-term effect of reputation damage notwithstanding, developers of very high quality or very low quality apps, stand to benefit from such promotions. However, the implications are not conclusive for the app developers in the center of this spectrum.

For appstores, long-term success depends on the satisfaction of both customers as well as app developers. There needs to be a complex trade-off between providing users with quality apps at low price, while at the same time mitigating potential losses to app developers' reputation and profits. Existing incentives to provide higher app visibility is primarily attractive to bottom ranked, low quality apps. Such practices may yield gains in market share for a short-term, but inhibit long-term customer retention.

Last, but not least, we find that increased app visibility on account of promotion increases brand awareness due to social media and word-of-mouth engagements. This effect not only drives future sales on the primary appstore i.e., Amazon Appstore, but also spills over across the markets onto other appstores like Google Playstore. This adds an additional complexity in measuring the *true* impact of promotions on the revenues of app developers. Indeed, a rational app developer should weigh the incremental revenue from across different appstores against short-term reputation damages, while assessing the merits of promotion. From the perspective of appstores, it seems odd that they would incur marketing costs to improve their market share while at the same time helping direct competitors, but as the Amazon *Free App of the Day* promotion example shows, Amazon likely prioritizes direct benefits to their own platform over indirect benefits to their competitors, Google Playstore [18].

Our findings thus contribute to both the academic literature and practitioners in the mobile app market in several important ways. Our study makes contribution to the growing body of research that utilizes publicly available e-commerce data to empirically validate research questions. It extends the existing knowledge about promotion strategies on emerging mobile app market, while also validating existing theories about consumer behaviors during discounted promotions. For the mobile app developers, our study provides important guidance about most influential factors in determining the success of marketing strategies. Because the smartphone appstores are almost certain to continue becoming even more competitive, the implications of understanding these marketing strategies are likely to become increasingly important.

References

Anderson C (2006) *The long tail: Why the future of business is selling less of more* (Hachette Books).

Askalidis G (2015) The Impact of Large Scale Promotions on the Sales and Ratings of Mobile Apps: Evidence from Apple's App store. *arXiv preprint arXiv:1506.06857* .

Barnes S (2016) Understanding Virtual Reality in Marketing: Nature, Implications and Potential .

Bart Y, Stephen AT, Sarvary M (2014) Which products are best suited to mobile advertising? a field study of mobile display advertising effects on consumer attitudes and intentions. *Journal of Marketing Research* 51(3):270–285.

[18] In late 2015, Amazon shut down the *Free App of the Day* promotion and replaced it with *Amazon Underground* which necessitated that a customer downloads the app from their platform to take advantage of promotion. They started marketing the *Amazon Underground* app instead of individual promoted apps.

Bartels B (2008) Beyond" fixed versus random effects": a framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. *The Society for Political Methodology* 1–43.

Bell A, Jones K (2015) Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* 3(01):133–153.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *The Quarterly journal of economics* 119(1):249–275.

Bryk AS, Raudenbush SW (1987) Application of hierarchical linear models to assessing change. *Psychological Bulletin* 101(1):147.

Brynjolfsson E, Hu Y, Smith MD (2003) Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* 49(11):1580–1596.

Brynjolfsson E, Hu YJ, Smith MD (2006) From niches to riches: Anatomy of the long tail .

Byers JW, Mitzenmacher M, Zervas G (2012) The Groupon effect on Yelp ratings: a root cause analysis. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 248–265 (ACM).

Chang K, Siddarth S, Weinberg CB (1999) The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science* 18(2):178–192.

Cheng HK, Tang QC (2010) Free trial or no free trial: Optimal software product design with network effects. *European Journal of Operational Research* 205(2):437–447.

Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon. com and Barnesandnoble. com. *Quantitative marketing and Economics* 1(2):203–222.

Chia PH, Yamamoto Y, Asokan N (2012) Is this app safe?: a large scale study on application permissions and risk signals. *Proceedings of the 21st International Conference on World Wide Web*, 311–320 (ACM).

Danaher P, Smith M, Ranasinghe K, Dagger T (2011) Assessing the effectiveness of mobile phone promotions. Technical report, working paper, Monash University.

Decker R, Trusov M (2010) Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing* 27(4):293–307.

Demidenko E (2013) *Mixed models: theory and applications with R* (John Wiley & Sons).

Edelman B, Jaffe S, Kominers SD (2016) To Groupon or not to Groupon: The profitability of deep discounts. *Marketing Letters* 27(1):39–53.

Engel JF, Blackwell RD, Miniard PW (1995) Consumer Behavior, 8th. *New York: Dryder* .

Erdem T, Sun B (2002) An empirical investigation of the spillover effects of advertising and sales promotions in umbrella branding. *Journal of Marketing Research* 39(4):408–420.

Garg R, Telang R (2012) Inferring app demand from publicly available data. *MIS Quarterly, Forthcoming* .

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

Ghose A, Ipeirotis P, Li B (2009) The economic impact of user-generated content on the internet: Combining text mining with demand estimation in the hotel industry. *Proceedings of the 20th Workshop on Information Systems and Economics (WISE).*

Ghose A, Sundararajan A (2006) Evaluating pricing strategy using e-commerce data: Evidence and estimation challenges. *Statistical Science* 131–142.

Kotler P, Keller KL (2006) Marketing Management 12e. *New Jersey* .

Leyland AH (2010) No quick fix: understanding the difference between fixed and random effect models. *Journal of Epidemiology & Community Health* 64(12):1027–1028.

Liu CZ, Au YA, Choi HS (2012) An empirical study of the freemium strategy for mobile apps: Evidence from the Google Play Market. *Proceedings of 33rd International Conference on Information Systems.*

Luca M (2011) Reviews, reputation, and revenue: The case of Yelp. com. *Com (September 16, 2011). Harvard Business School NOM Unit Working Paper* (12-016).

Moe WW (2003) Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology* 13(1-2):29–39.

Rogosa DR, Willett JB (1985) Understanding correlates of change by modeling individual differences in growth. *Psychometrika* 50(2):203–228.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Shankar V, Balasubramanian S (2009) Mobile marketing: a synthesis and prognosis. *Journal of interactive marketing* 23(2):118–129.

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* 74(2):133–148.

Table 1 Summary Statistics of Amazon Appstore.

	Treatment		Control		Overall	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Price (USD)$^{(L)}$	2.31	1.45	2.58	3.08	2.57	3.05
File Size (megabytes)$^{(L)}$	44.28	66.10	35.72	78.70	35.93	78.42
Description Length (characters)$^{(L)}$	6.79	0.78	6.81	0.79	6.81	0.79
Number of Permissions	4.32	3.07	4.64	3.74	4.63	3.72
Number of Screenshots	7.23	2.64	6.62	2.87	6.64	2.87
Maturity Rating						
All Ages	0.73	0.44	0.75	0.43	0.75	0.43
Guidance Suggested	0.27	0.44	0.23	0.42	0.23	0.42
Mature	0.00	0.00	0.02	0.14	0.02	0.14
Adult	0.00	0.00	0.00	0.02	0.00	0.02
In-App Purchase (1:yes, 0:no)	0.27	0.44	0.14	0.34	0.14	0.35
Min. Android Version	2.08	0.57	2.09	0.72	2.09	0.72
App Age (months)	26.61	17.06	29.12	17.07	29.06	17.08
Version	1.60	1.32	1.79	2.59	1.79	2.56
Apps By Developer$^{(L)}$	1.67	1.32	1.96	1.62	1.96	1.62
Recommendation Count$^{(L)}$	1.96	1.37	1.36	1.08	1.37	1.09
Category						
Education	0.05	0.22	0.07	0.26	0.07	0.26
Games	0.72	0.45	0.40	0.49	0.41	0.49
Kids	0.05	0.22	0.09	0.28	0.09	0.28
Music & Audio	0.02	0.15	0.04	0.20	0.04	0.20
Photo & Video	0.02	0.16	0.02	0.15	0.02	0.15
Productivity	0.05	0.23	0.07	0.26	0.07	0.26
Other	0.08	0.27	0.30	0.46	0.29	0.45
User Review Count$^{(L)}$	5.44	1.41	2.86	1.49	2.93	1.55
User Rating	3.97	0.50	3.66	0.93	3.67	0.92
Observations	1619		62545		64164	

Note: The sample period is from February, 2015, to December, 2015.
$^{(L)}$ denotes Logarithm of the variable.

Table 2 Summary Statistics of Google Playstore for Amazon FAD promoted apps.

	Mean	Std. dev.
Price (USD)$^{(L)}$	1.054	0.611
File Size (megabytes)$^{(L)}$	3.385	0.998
Description Length (characters)$^{(L)}$	7.203	0.668
Number of Screenshots$^{(L)}$	2.561	0.471
Maturity Rating		
All Ages	0.716	0.451
Guidance Suggested	0.090	0.287
Mature	0.135	0.341
Adult	0.059	0.236
In-App Purchase (1:yes, 0:no)	0.255	0.436
Min. Android Version	2.066	0.439
Number of Versions$^{(L)}$	0.860	0.360
User Review Count$^{(L)}$	4.626	2.031
User Rating	4.266	0.514
Observations	65952	

Note: $^{(L)}$ denotes Logarithm of the variable.

Table 3 Rank Categories

	Sales Rank
Rank Category 1	1 - 1984
Rank Category 2	1984 - 4573
Rank Category 3	4574 - 22189

Table 4 Impact of FAD promotion

	(1) Sales Rank$^{(L)}$		(2) Monthly Review Count$^{(L)}$		(3) User Rating	
Mean Effects:						
After	−0.282***	(−3.74)	2.931***	(33.22)	−0.160***	(−5.71)
Age	0.027***	(39.64)	−0.048***	(−66.87)	−0.008***	(−14.36)
Post	0.226***	(7.83)	−1.088***	(−35.96)	−0.010***	(−2.76)
Post^2	−0.020***	(−4.72)	0.113***	(24.51)		
Effect of App Characteristics:						
In-App Purchase	−0.123***	(−3.65)	0.066***	(4.01)	0.108***	(3.71)
Maturity Rating						
Guidance Suggested	−0.008	(−0.33)	0.002	(0.17)	−0.114***	(−4.34)
Mature	0.189***	(3.44)	−0.061***	(−2.81)	−0.130*	(−1.96)
Adult	0.916***	(11.00)	−0.162***	(−3.82)	−1.250*	(−1.92)
Price$^{(L)}$	0.019	(1.21)	0.060***	(8.25)	0.120***	(6.79)
Size$^{(L)}$	−0.115***	(−13.69)	0.026***	(6.82)	0.048***	(5.11)
Number of Permissions$^{(L)}$	0.018	(1.05)	0.050***	(6.32)	−0.177***	(−9.40)
Number of Screenshots$^{(L)}$	−0.145***	(−4.47)	0.090***	(6.58)	0.227***	(6.60)
Description Length$^{(L)}$	−0.083***	(−6.36)	0.032***	(5.81)	0.114***	(7.66)
Apps By Developer$^{(L)}$	0.023***	(3.41)	−0.029***	(−10.60)	0.050***	(6.56)
Number of Versions$^{(L)}$	−0.053**	(−2.28)	0.059***	(5.26)	0.025	(0.90)
Min. Android Version	−0.085***	(−4.84)	0.058***	(7.02)	0.087***	(5.12)
Constant	8.660***	(364.81)	0.518***	(48.85)	3.715***	(144.39)
Between Effects:						
Age *(between)*	0.008***	(10.76)	−0.005***	(−15.80)	−0.005***	(−6.06)
After *(between)*	−20.001	(−1.01)	24.181**	(2.26)	−1.559	(−1.45)
Post *(between)*	14.326	(0.90)	−16.930**	(−2.00)	0.541	(1.50)
Post^2 *(between)*	−1.849	(−0.86)	2.216**	(1.96)		
Interaction Effects:						
After × In-App Purchase	−0.083	(−0.61)	0.033	(0.22)	0.020	(0.26)
After × Price$^{(L)}$	−0.068	(−0.66)	0.002	(0.02)	0.044	(0.92)
After × Size$^{(L)}$	0.142***	(2.94)	−0.100*	(−1.73)	−0.023	(−0.64)
After × Number of Permissions$^{(L)}$	0.187	(1.43)	−0.254	(−1.57)	−0.030	(−0.54)
After × Number of Screenshots$^{(L)}$	−0.327**	(−2.17)	0.037	(0.17)	−0.040	(−0.37)
After × Description Length$^{(L)}$	−0.196***	(−2.69)	0.066	(0.73)	0.024	(0.52)
After × Apps By Developer$^{(L)}$	0.048	(0.61)	0.046	(0.73)	−0.026	(−0.77)
After × Number of Versions$^{(L)}$	0.043	(0.31)	−0.143	(−0.80)	0.050	(0.54)
After × Min. Android Version	0.007	(0.10)	−0.085	(−1.06)	0.044	(1.25)
Variance Components:						
var(Age)	0.001***	(0.00)	0.000***	(0.00)	0.002***	(0.00)
var(Constant)	1.836***	(0.06)	0.427***	(0.02)	1.399***	(0.04)
corr(Age, Constant)	−0.035***	(0.00)	−0.008***	(0.00)	−0.030***	(0.00)
var(Residual)	0.086***	(0.00)	0.154***	(0.00)	0.026***	(0.00)
Observations	64164		64164		63735	
AIC	75444		82666		24077	
BIC	75753		82974		24367	
Pseudo Log Likelihood	-37688		-41299		-12006	

Note: Referent level for maturity rating is 'All Ages'. $^{(L)}$ denotes Logarithm of the variable. Cluster-robust t-statistics (at app level) are shown in parentheses. Truncated version of the table due to space constraints.

Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 5 Heterogeneous impact of FAD promotion

	(1) Sales Rank$^{(L)}$		(2) Monthly Review Count$^{(L)}$		(3) User Rating	
Mean Effects:						
After	−0.175**	(−2.55)	2.304***	(19.34)	−0.082***	(−2.68)
2^{nd} Rank Category × After	−0.080	(−0.59)	0.895***	(5.11)	−0.112**	(−2.08)
3^{rd} Rank Category × After	−0.366***	(−3.34)	0.951***	(5.70)	−0.147**	(−2.00)
Age	0.027***	(9.40)	−0.062***	(−24.48)	−0.009***	(−4.87)
2^{nd} Rank Category × Age	0.005	(1.49)	0.004	(1.33)	−0.001	(−0.37)
3^{rd} Rank Category × Age	−0.002	(−0.75)	0.019***	(7.08)	0.001	(0.72)
Post	0.140***	(3.92)	−0.884***	(−16.94)	−0.010**	(−2.51)
2^{nd} Rank Category × Post	0.167***	(2.60)	−0.323***	(−4.44)	0.000	(0.06)
3^{rd} Rank Category × Post	0.070	(1.05)	−0.279***	(−4.03)	0.003	(0.33)
Post^2	−0.014***	(−2.59)	0.094***	(12.43)		
2^{nd} Rank Category × Post^2	−0.016	(−1.64)	0.037***	(3.32)		
3^{rd} Rank Category × Post^2	−0.001	(−0.14)	0.022**	(2.13)		
2^{nd} Rank Category	1.452***	(51.23)	−0.564***	(−21.38)	−0.083*	(−1.93)
3^{rd} Rank Category	2.721***	(100.78)	−0.889***	(−37.13)	−0.126***	(−3.55)
Constant	6.471***	(237.89)	1.249***	(53.35)	3.842***	(111.14)
Between Effects:						
Age *(between)*	0.003***	(10.08)	−0.004***	(−18.24)	−0.007***	(−9.28)
After *(between)*	−23.040*	(−1.88)	24.743***	(2.90)	−1.652*	(−1.68)
Post *(between)*	17.894*	(1.88)	−17.817***	(−2.65)	0.579*	(1.74)
Post^2 *(between)*	−2.376*	(−1.90)	2.354***	(2.62)		
Variance Components:						
var(Age)	0.000***	(0.000)	0.000***	(0.000)	0.002***	(0.000)
var(Constant)	0.518***	(0.024)	0.291***	(0.016)	1.448***	(0.360)
corr(Age, Constant)	−0.011***	(0.001)	−0.005***	(0.000)	−0.030***	(0.001)
var(Residual)	0.096***	(0.027)	0.153***	(0.002)	0.026***	(0.001)
Observations	64164		64164		63735	
AIC	61793		78901		24558	
BIC	62002		79110		24730	
Pseudo Log Likelihood	-30874		-39428		-12260	

Note: Referent level for maturity rating is 'All Ages'. $^{(L)}$ denotes Logarithm of the variable. Cluster-robust t-statistics (at app level) are shown in parentheses. Truncated version of the table due to space constraints.

Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 6 Comparison between multilevel models and fixed effects models

	(1) Sales Rank$^{(L)}$		(2) Sales Rank$^{(L)}$	
Mean Effects:				
After	−0.309***	(−4.04)	−0.309***	(−4.04)
Age	0.027***	(39.28)	0.027***	(39.27)
Post	0.220***	(7.74)	0.220***	(7.74)
Post^2	−0.020***	(−4.67)	−0.020***	(−4.67)
Effect of App Characteristics:				
In-App Purchase	−0.147***	(−4.40)		
Maturity Rating				
Guidance Suggested	−0.018	(−0.77)		
Mature	0.183***	(3.46)		
Adult	0.845***	(8.71)		
Price$^{(L)}$	0.007	(0.45)		
Size$^{(L)}$	−0.124***	(−15.52)		
Number of Permissions$^{(L)}$	0.003	(0.18)		
Number of Screenshots$^{(L)}$	−0.185***	(−6.28)		
Description Length$^{(L)}$	−0.074***	(−5.84)		
Apps By Developer$^{(L)}$	0.026***	(4.18)		
Number of Versions$^{(L)}$	−0.044**	(−2.02)		
Min. Android Version	−0.072***	(−4.43)		
Constant	8.591***	(363.50)	7.781***	(389.23)
Between Effects:				
After*(between)*	−14.107	(−0.82)		
Age*(between)*	0.009***	(13.31)		
Post*(between)*	9.461	(0.69)		
Post^2*(between)*	−1.199	(−0.66)		
Interaction Effects:				
After × In-App Purchase	−0.077	(−0.52)	−0.077	(−0.52)
After × Price$^{(L)}$	−0.058	(−0.52)	−0.058	(−0.52)
After × Size$^{(L)}$	0.149***	(3.05)	0.149***	(3.05)
After × Number of Permissions$^{(L)}$	0.130	(0.93)	0.130	(0.93)
After × Number of Screenshots$^{(L)}$	−0.319**	(−2.01)	−0.319**	(−2.01)
After × Description Length$^{(L)}$	−0.182**	(−2.37)	−0.182**	(−2.37)
After × Apps By Developer$^{(L)}$	0.061	(0.80)	0.061	(0.80)
After × Number of Versions$^{(L)}$	−0.039	(−0.27)	−0.039	(−0.27)
After × Min. Android Version	−0.019	(−0.24)	−0.019	(−0.24)
Variance Components:				
var(Constant)	0.903***	(0.02)		
var(Residual)	0.105***	(0.00)		
Observations	64164		64164	
AIC	79533		24926	
BIC	79823		25044	
Pseudo Log Likelihood	-39734		-12450	

Note: Referent level for maturity rating is 'All Ages'. $^{(L)}$ denotes Logarithm of the variable. Cluster-robust t-statistics (at app level) are shown in parentheses. Truncated version of the table due to space constraints.

Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 7 Summary Statistics of Amazon Appstore (matched sample)

	Treatment		Control		Overall	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Price (USD)$^{(L)}$	2.31	1.45	2.51	2.47	2.49	2.39
File Size (megabytes)$^{(L)}$	44.28	66.10	47.87	84.67	47.50	82.97
Description Length (characters)$^{(L)}$	6.79	0.78	6.85	0.80	6.85	0.80
Number of Permissions	4.32	3.07	4.73	3.72	4.69	3.66
Number of Screenshots	7.24	2.63	7.24	2.98	7.24	2.95
Maturity Rating						
All Ages	0.73	0.44	0.75	0.43	0.75	0.43
Guidance Suggested	0.27	0.44	0.23	0.42	0.24	0.43
Mature	0.00	0.00	0.01	0.12	0.01	0.11
In-App Purchase (1:yes, 0:no)	0.27	0.44	0.23	0.42	0.23	0.42
Min. Android Version	2.08	0.57	2.08	0.63	2.08	0.62
App Age (months)	26.61	17.06	27.13	14.66	27.07	14.92
Version	1.60	1.32	1.67	1.88	1.67	1.83
Apps By Developer$^{(L)}$	1.67	1.32	2.07	1.61	2.02	1.59
Recommendation Count$^{(L)}$	1.96	1.37	1.74	1.14	1.76	1.17
Category						
Education	0.05	0.22	0.06	0.24	0.06	0.24
Games	0.72	0.45	0.56	0.50	0.58	0.49
Kids	0.05	0.22	0.10	0.29	0.09	0.29
Music & Audio	0.02	0.15	0.03	0.16	0.03	0.16
Photo & Video	0.02	0.16	0.02	0.14	0.02	0.14
Productivity	0.05	0.23	0.06	0.23	0.06	0.23
Other	0.08	0.27	0.18	0.38	0.17	0.37
User Review Count$^{(L)}$	5.44	1.41	4.29	1.41	4.41	1.45
User Rating	3.97	0.50	3.96	0.61	3.96	0.60
Observations	1619		14206		15825	

Note: The sample period is from February, 2015, to December, 2015.
$^{(L)}$ denotes Logarithm of the variable.

Table 8 Impact of FAD promotion (matched sample)

	(1) Sales Rank$^{(L)}$		(2) Monthly Review Count$^{(L)}$		(3) User Rating	
Mean Effects:						
After	−0.253***	(−3.51)	2.906***	(33.31)	−0.172***	(−5.50)
Age	0.030***	(17.87)	−0.055***	(−32.80)	−0.008***	(−8.49)
Post	0.220***	(7.63)	−1.080***	(−35.68)	−0.009**	(−2.44)
Post^2	−0.020***	(−4.70)	0.113***	(24.51)		
Effect of App Characteristics:						
In-App Purchase	0.037	(0.46)	−0.010	(−0.26)	0.023	(0.60)
Maturity Rating						
Guidance Suggested	−0.017	(−0.23)	0.036	(0.97)	0.011	(0.29)
Mature	0.048	(0.21)	−0.031	(−0.32)	0.092	(0.87)
Price$^{(L)}$	0.007	(0.13)	0.137***	(4.96)	0.084***	(2.94)
Size$^{(L)}$	−0.154***	(−5.37)	0.042***	(3.02)	0.036**	(2.20)
Number of Permissions$^{(L)}$	0.009	(0.15)	0.113***	(4.23)	−0.130***	(−4.21)
Number of Screenshots$^{(L)}$	−0.335***	(−3.24)	0.110**	(2.29)	0.103*	(1.84)
Description Length$^{(L)}$	−0.285***	(−7.39)	0.092***	(5.22)	0.069***	(2.95)
Apps By Developer$^{(L)}$	0.022	(0.90)	−0.051***	(−5.11)	0.016	(1.26)
Number of Versions$^{(L)}$	−0.106	(−1.27)	0.151***	(3.66)	−0.020	(−0.38)
Min. Android Version	−0.256***	(−4.67)	0.176***	(6.61)	0.044	(1.64)
Constant	8.168***	(111.63)	0.763***	(21.70)	4.174***	(94.54)
Between Effects:						
Age *(between)*	−0.001	(−0.30)	−0.003**	(−2.29)	−0.012***	(−8.26)
After *(between)*	−5.703	(−0.30)	18.510*	(1.82)	−2.076**	(−2.56)
Post *(between)*	4.193	(0.27)	−12.967	(−1.60)	0.607**	(2.20)
Post^2 *(between)*	−0.544	(−0.26)	1.710	(1.58)		
Interaction Effects:						
After × In-App Purchase	−0.075	(−0.59)	0.035	(0.23)	0.015	(0.21)
After × Price$^{(L)}$	−0.062	(−0.64)	−0.006	(−0.05)	0.033	(0.72)
After × Size$^{(L)}$	0.142***	(3.00)	−0.103*	(−1.76)	−0.020	(−0.60)
After × Number of Permissions$^{(L)}$	0.175	(1.40)	−0.254	(−1.56)	−0.023	(−0.46)
After × Number of Screenshots$^{(L)}$	−0.352**	(−2.35)	0.033	(0.15)	−0.029	(−0.29)
After × Description Length$^{(L)}$	−0.191***	(−2.74)	0.068	(0.74)	0.022	(0.49)
After × Apps By Developer$^{(L)}$	0.038	(0.50)	0.046	(0.74)	−0.026	(−0.80)
After × Number of Versions$^{(L)}$	0.076	(0.56)	−0.168	(−0.93)	0.045	(0.55)
After × Min. Android Version	0.037	(0.55)	−0.091	(−1.15)	0.037	(1.15)
Variance Components:						
var(Age)	0.003***	(0.00)	0.000***	(0.00)	0.001***	(0.00)
var(Constant)	2.751***	(0.18)	0.812**	(0.08)	0.462***	(0.03)
corr(Age, Constant)	−0.057***	(0.00)	−0.016***	(0.00)	−0.009***	(0.00)
var(Residual)	0.102***	(0.01)	0.225***	(0.00)	0.010***	(0.00)
Observations	15825		15825		15823	
AIC	20126		26457		-11712	
BIC	20379		26711		-11474	
Pseudo Log Likelihood	-10030		-13196		5887	

Note: Referent level for maturity rating is 'All Ages'. $^{(L)}$ denotes Logarithm of the variable. Cluster-robust t-statistics (at app level) are shown in parentheses. Truncated version of the table due to space constraints.

Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 9 Heterogeneous impact of FAD promotion (matched sample)

	(1) Sales Rank$^{(L)}$		(2) Monthly Review Count$^{(L)}$		(3) User Rating	
Mean Effects:						
After	−0.154**	(−2.21)	2.292***	(19.13)	−0.089***	(−2.94)
2^{nd} Rank Category × After	−0.121	(−0.91)	0.920***	(5.22)	−0.099*	(−1.93)
3^{rd} Rank Category × After	−0.369***	(−3.33)	0.983***	(5.86)	−0.131*	(−1.90)
Age	0.023***	(5.82)	−0.059***	(−16.53)	−0.008***	(−5.91)
2^{nd} Rank Category × Age	0.016***	(2.88)	−0.003	(−0.65)	−0.003	(−1.17)
3^{rd} Rank Category × Age	0.006	(1.31)	0.009**	(2.13)	0.002	(0.94)
Post	0.150***	(4.24)	−0.888***	(−17.05)	−0.013***	(−3.63)
2^{nd} Rank Category × Post	0.145**	(2.25)	−0.314***	(−4.32)	0.006	(0.81)
3^{rd} Rank Category × Post	0.058	(0.87)	−0.266***	(−3.85)	0.009	(1.09)
Post^2	−0.015***	(−2.77)	0.094***	(12.47)		
2^{nd} Rank Category × Post^2	−0.015	(−1.54)	0.037***	(3.30)		
3^{rd} Rank Category × Post^2	−0.001	(−0.06)	0.022**	(2.10)		
2^{nd} Rank Category	1.612***	(29.05)	−0.605***	(−14.41)	−0.036	(−0.63)
3^{rd} Rank Category	2.676***	(52.73)	−0.979***	(−27.03)	−0.068	(−1.43)
Constant	6.357***	(108.93)	1.415***	(33.81)	4.251***	(98.22)
Between Effects:						
Age*(between)*	0.002*	(1.71)	−0.003***	(−4.05)	−0.013***	(−10.46)
After*(between)*	−18.751	(−1.54)	23.161***	(2.83)	−2.272***	(−3.11)
Post*(between)*	14.914	(1.54)	−16.878***	(−2.62)	0.675***	(2.71)
Post^2*(between)*	−1.998	(−1.53)	2.243***	(2.60)		
Variance Components:						
var(Age)	0.001***	(0.000)	0.000***	(0.000)	0.001***	(0.000)
var(Constant)	1.114***	(0.107)	0.604***	(0.062)	0.466***	(0.027)
corr(Age, Constant)	−0.032***	(0.003)	−0.011***	(0.001)	−0.349***	(0.001)
var(Residual)	0.110***	(0.006)	0.224***	(0.004)	0.010***	(0.008)
Observations	15825		15825		15823	
AIC	18218		25771		-11684	
BIC	18395		25947		-11538	
Pseudo Log Likelihood	-9086		-12862		5861	

Note: Referent level for maturity rating is 'All Ages'. $^{(L)}$ denotes Logarithm of the variable. Cluster-robust t-statistics (at app level) are shown in parentheses. Truncated version of the table due to space constraints.

Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

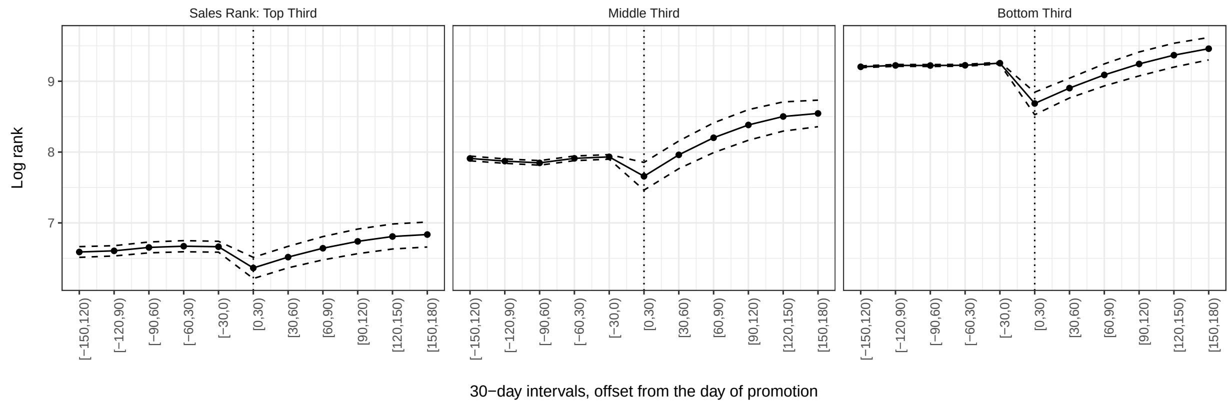


Figure 2 Heterogeneity of impact on sales rank

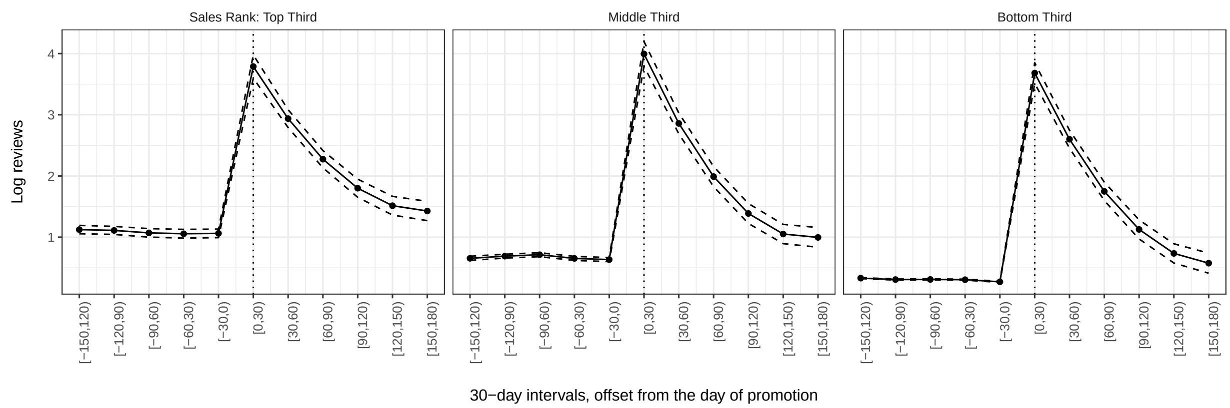


Figure 3 Heterogeneity of impact on number of reviews

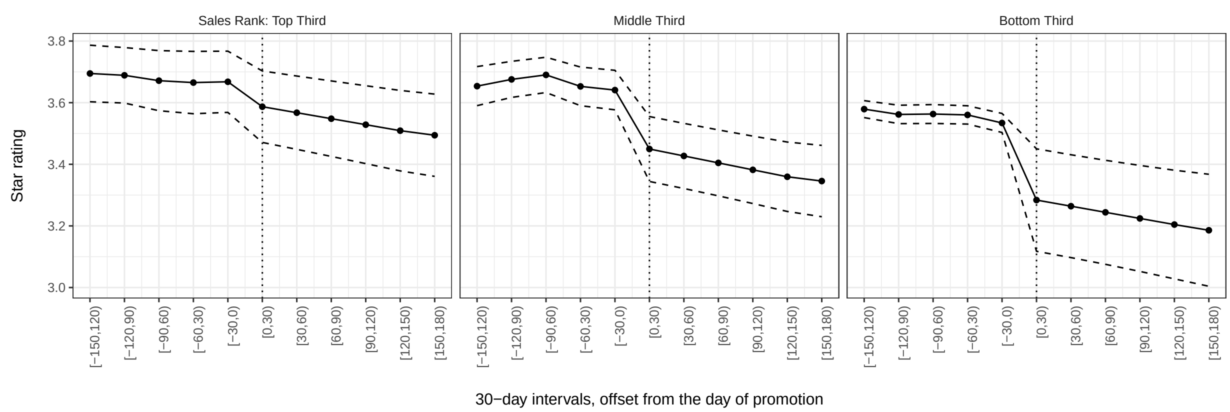


Figure 4 Heterogeneity of impact on star rating

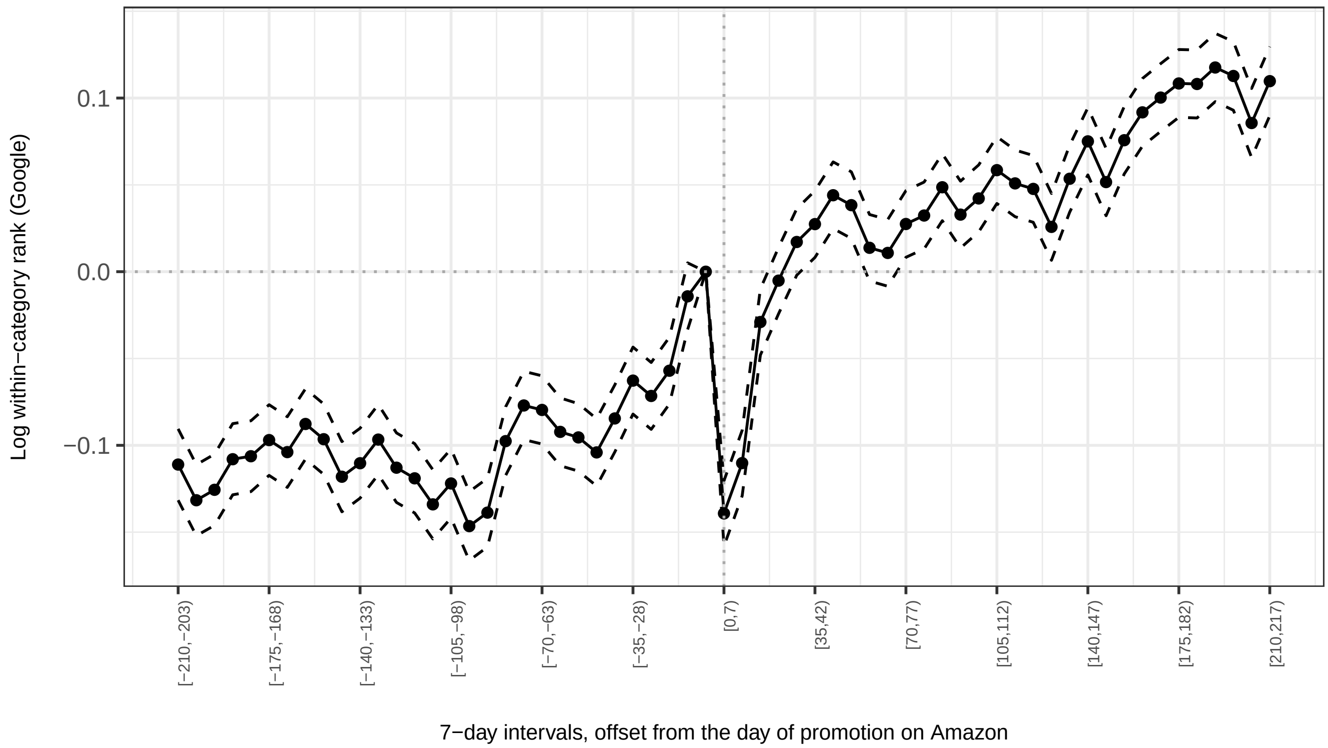


Figure 5 Cross-market spillover effect on sales rank

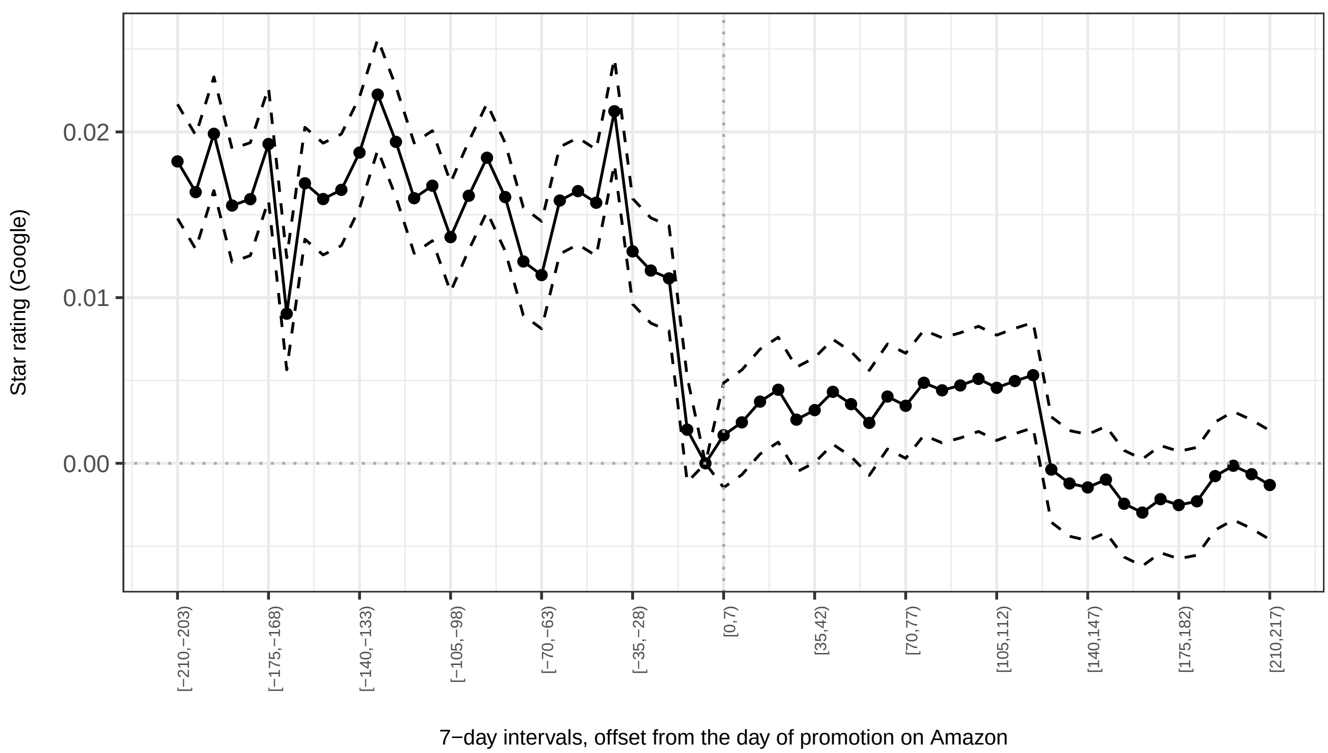


Figure 6 Cross-market spillover effect on star rating

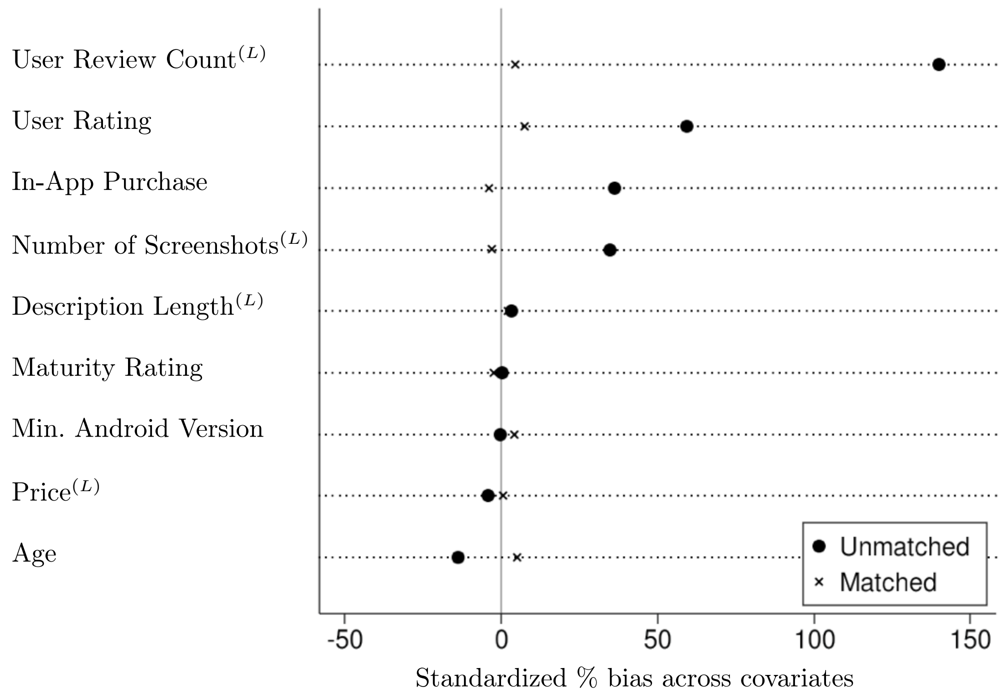


Figure 7 Matching the Treatment and Control Apps

8. Appendix

8.1. How app characteristics impact general trends on Amazon Appstore

In this section, we describe interpret the estimates in Table 4 and describe in detail the effects of time-invariant app characteristics on the general trends in Amazon Appstore. We find that a 10% increase in an app's price increases app's monthly reviews by 0.6% and average star rating by 0.01 stars. However, price does not affect an app's sales rank significantly. In-app purchase options tend to be common among the top apps. Apps with in-app purchase options have 12% better sales rank on an average, receive 6% more monthly reviews and have 0.1 star higher user rating. On the Amazon Appstore, longer waiting times to download sophisticated apps of larger size do not seem to adversely affect demand for an app. In fact, a 10% increase in app size improves sales rank by 1.2% and increases monthly reviews by 0.3% with an average 0.005 stars increase per user rating. One extra permission results in an average 0.03 stars decrease in user rating, while increasing monthly reviews by 1.2 %. Expectedly, one extra screenshot on the app profile page improves sales rank by 2% and increases monthly reviews by 1.3% at an average 0.03 stars more than usual. We find that 10% increase in description length results in 0.8% improvement in sales rank and an increase of 0.3% monthly reviews. Consumers tend to reward such an app with 0.01 stars more on an average. Thus, it is important for app developers to provide their potential customers with sufficient graphic and textual description for their apps. In terms of maturity rating (age restrictions), compared to "All Ages", apps belonging to "Guidance Suggested", "Mature" or "Adult" categories exhibit worse sales rank, lower number of monthly reviews and lower star ratings. Thus, references to violence, sex, drugs and alcohol have a negative impact on the app performance. We find that developers do not always benefit from publishing more apps on the Amazon Appstore. However, apps which are regularly updated tend to have better sales rank and get more reviews every month. Importantly, as the minimum Android version supported by the app increases, the quality of app improves too, due to evolved functionalities in higher version of Android. This results in an improvement in sales rank, an increase in monthly reviews and user ratings 0.09 stars higher than average.

We believe that a better understanding of how different app characteristics affect the performance of apps on Amazon Appstore will help the app developers not only in the marketing decisions, but also during the development cycle of an app.