Unintended Consequences of Platform Monetization on Digital Cultural Markets: Evidence from a Natural Experiment on Goodreads

Kai Zhu[1], Qiaoni Shi[2], and Shrabastee Banerjee*[3]

[1,2] *Bocconi University*
[3] *Tilburg University*

Abstract

How can a platform capture the value it creates for its users while fostering a diverse and inclusive marketplace? This study explores the potential consequences of monetizing a prominent promotional program within a platform market for book promotions, specifically the Giveaways program on Goodreads.com. Participation in this program was free until January 2018, when Goodreads introduced a policy that imposed a flat fee on authors or publishers who wish to promote their books in this program. We employ large-scale, fine-grained data to examine responses from both the supply side and the demand side of the two-sided market to this monetization strategy. Our results suggest that this policy results in a more concentrated book supply in the promotional market, increasing the market share of major and established publishers and authors at the expense of smaller entities with fewer resources. The policy also reduces book diversity in the marketplace, with a few popular genres becoming more dominant while niche genres diminish further. These supply-side changes ultimately impact consumers, as the policy exacerbates the effect of price promotion on word of mouth (i.e., increased review volume but lower valence). An analysis of textual reviews and rating dispersion reveals that consumer-book mismatch drives this consumer response. Our findings carry policy implications for platform owners and policymakers: to establish and maintain a healthy ecosystem, platforms need to mitigate potential undesirable effects from monetization strategies by implementing more flexible and nuanced incentive structures for various participants.

Keywords: Platform Monetization, Cultural Products, Online Marketplace, Two-sided Market

Acknowledgements: The authors would like to thank seminar participants at EMAC, SCECR, WISE, the Digital Economy Workshop, the Paris Conference on Digital Economics, the Marketing department faculty at London Business School, University College London, Imperial College London, Bocconi University, Singapore Management University, Indian Institute of Management Ahmedabad , and specifically Christian Peukert, Michail Batikas, Dominik Gutt, Ananya Sen, Michelle Andrews, Xiao Liu, Yiting Deng and Davide Proserpio for their helpful feedback.

*Note: All authors contributed equally to this project.

1 Introduction

Over the past decade, the internet and digitization have drastically transformed the markets for cultural products, such as books, music, movies, and television shows. With the advent of search engines, recommendation systems, and user-generated content, new technologies have revolutionized the way these cultural products are produced, distributed, promoted, and consumed (Waldfogel, 2017; Peukert, 2019). In particular, digital platforms now play a vital role in our engagement with these markets. In recent years, various platform businesses for cultural goods have emerged, including content streaming platforms like Netflix and Spotify, online retail marketplaces such as Amazon Books, and social catalog platforms like Goodreads and Douban. These platforms have fundamentally altered how we interact with cultural products, disrupting traditional industry business models.

Digital technologies have led to a significant reduction in content production and distribution costs, ultimately resulting in lower costs of cultural participation for individuals (Peukert, 2019). This supply-side implication has caused an explosion of new products in these markets (Waldfogel, 2017). Lower barriers to entry in various cultural fields now enable anyone to become a "creator", even without professional training. Individuals can write their own book, compose a song, or shoot a short video, and share it with a broad audience by publishing on various digital distribution platforms such as Amazon Kindle Direct Publishing, SoundCloud, or YouTube. As a result, digitization has created a "long tail" of diverse content catering to the niche preferences of different individuals who were previously underserved.

However, the abundance of new content has lead to greater competition among content creators, necessitating greater self-promotion and marketing efforts to stand out in a crowded marketplace (Nagaraj and Ranganathan, 2022). This is particularly relevant in the book publishing industry, which is the focus of our research. With over four million new titles published in the U.S. in 2021, including self-published and commercially published books, content creators need to have the necessary resources and channels to effectively promote their work to consumers who are inundated with new products. The number of new titles has been growing steadily in recent years, emphasizing the need for effective promotion strategies in a highly competitive market.

In the competitive landscape of cultural products, platform decisions about product promotion significantly impact consumer discovery and product success (Aguiar and Waldfogel, 2021). A growing body of evidence suggests that consumption patterns of digital goods are becoming more concentrated and increasingly reliant on platforms (Fleder and Hosanagar, 2009; Nielsen and Fletcher, 2022; Guinaudeau et al., 2022). In the context of book publishing—a market already characterized by high concentration—platform decisions to promote or incentivize independent or niche books and authors are essential for achieving creators' self-promotion goals. These factors raise concerns about how platforms exercise market power and the potential implications for the overall marketplace. The issue of platform power is especially relevant in the realm of cultural products, where fostering a diverse and inclusive environment for content creation is critical. This environment is not only important for equity reasons but also for ensuring product variety and diverse representation, catering to horizontally differentiated consumer preferences in cultural products.

On the other hand, platform owners must establish effective policies to manage value creation within their ecosystems, as the sustainable growth of a platform business depends on sound governance. The concept of platform governance stems from the idea that platforms serve as overseers of a micro-economy and are, therefore, responsible for formulating policies and mechanisms that ensure favorable outcomes for the entire ecosystem (Foerderer et al., 2021). Given the control platform owners exercise and their ability to interact with participants across diverse market segments, they are uniquely positioned to develop mechanisms for achieving these objectives (Rietveld et al., 2019).

Among the governance decisions a platform must tackle, formulating a robust monetization strategy is one of the most crucial (Farronato, 2017; Parker et al., 2016). To maintain their operations, all platform businesses need to develop methods for generating revenue streams by designing strategies to charge users for the services provided. This complex issue encompasses several key decisions for platform owners, including determining how to charge (e.g., subscription fee, transaction fee, enhanced access fee) and whom to charge (e.g., one or both sides of the market, offering a freemium model for price-sensitive consumers, charging full price overall but subsidizing stars). As highlighted in Parker et al. (2016), "monetization is one of the most challenging and intriguing issues that any platform company must address."

Despite its importance, monetization remains an insufficiently understood aspect of platform governance. This is particularly true for platforms focused on cultural products, such as books, movies, and music, where a suboptimal monetization strategy may lead to an unfavorable equilibrium by pricing out underrepresented participants and negatively impacting product diversity.

In this paper, we aim to deepen our understanding of platform monetization through an empirical examination of a case featuring a exogenous policy change: the introduction of participation fees in a digital book promotion marketplace. Our study focuses on Goodreads.com, a social cataloging platform that allows users to search for book information, create reading lists, write reviews, and engage with friends about reading activities. Known for its role in book discovery and its extensive repository of user-generated reviews, Goodreads has attracted over 125 million registered users as of 2021. We specifically examine the monetization of Goodreads' Giveaways program, a promotional tool directly connected to both book discovery and review generation. Essentially, authors and publishers can launch campaigns to 'give away' their books, increasing awareness, generating reader interest, and garnering reviews. Over recent years, the Giveaways program has emerged as an immensely popular means of enhancing a book's exposure and helping users find their next read on the platform.

We leverage a natural experiment that occurred when Goodreads began monetizing this program. In January 2018, the platform introduced a flat fee (starting at $119) for authors and publishers to give away their works. The sudden announcement provided only about two months for platform participants to adapt, making it a relatively clean natural experiment suitable for causal analysis. Another benefit of our research setting is the ability to observe the dynamics of both the supply side (authors and publishers) and the demand side (readers) in the two-sided book promotion market. This enables a comprehensive exploration of various factors within the platform ecosystem.

Although the new policy generates a revenue stream for the platform, the monetization mechanism raises concerns about potentially damaging digital cultural participation, especially for independent authors and small publishers. This could negatively impact consumers by diminishing the long tail and decreasing product variety[1]. Conversely, one might argue that monetization filters out low-quality participants from the supply side, thereby increasing consumer satisfaction. Driven by these opposing predictions, we investigate the effects of Goodreads' specific governance decision to monetize its promotional feature by imposing a flat fee on the supply side on its marketplace.

Utilizing comprehensive and detailed data from (1) the Giveaways program; (2) book, author, and publisher information; and (3) consumer reviews and ratings, we uncover several novel insights regarding platform monetization within the context of Goodreads. First, we observe a significant overall decline in marketplace

[1]Such concerns have also been voiced by the broader author community, for example, at https://www.theverge.com/2017/11/29/16714972/goodreads-giveaways-program-changing-standard-premium-tiers-authors

engagement following the implementation of the monetization policy, demonstrated by both the reduced number of Giveaways campaigns hosted on the platform and the diminished reader crowd requesting books. This substantial decrease in participation (approximately 66%) adversely affects word of mouth both on and off the platform (e.g., Twitter mentions and retweets). Moreover, supply-side monetization through a flat price does not impact all participants uniformly. While all entities experience a decrease in their offerings in absolute terms, large publishers gain market share in the Giveaways program in relative percentage terms. Conversely, small publishers lose market share, and self-published print books are almost pushed out of the market. Notably, less estalished authors' and female authors' participation in the program declines disproportionately following monetization. By examining the concentration of publishers participating in Giveaways over time using the Hirfindahl-Hirschman Index (HHI), we also identify a sharp increase in supply concentration among publishers (2.5 times relative to baseline).

Second, we discover that these supplier-related effects contribute to a shift in product characteristics within the marketplace. Specifically, the diversity of book genres experiences a sudden drop around the time of the monetization policy's enactment. Delving further into the proportions of different genres, we observe a "rich get richer, poor get poorer" phenomenon in play—popular genres increase their proportions, while underrepresented genres further diminish due to the pricing mechanism.

Lastly, we focus on the demand side and investigate consumer reactions to the supply-side changes that occur after monetization. Employing an event-study design and various difference-in-differences methods and specifications (Callaway and Sant'Anna, 2021), we observe that following the Giveaways program monetization, books experience a decline in average ratings while review volume rises, consistent with literature on price promotion (e.g., the "Groupon effect" identified by Byers et al., 2012b). More significantly, the monetization of the Giveaways program further exacerbates this effect—compared to books participating in the program prior to monetization, average ratings decrease by approximately 0.1 stars and review volume increases by 10 for books participating post-monetization. We hypothesize that the primary mechanism driving these results is the deterioration of consumer-product matches after the policy change, due to increased supply concentration and reduced product variety in the market.

To support this hypothesis, we conduct a natural language processing (NLP) analysis of review texts to directly infer signs of mismatch. We ask MTurkers to rate whether a negative review stems from fit- or quality-related issues and collect labels for 3,000 reviews. Using this set of manually labeled reviews as training data, we fine-tune a pre-trained BERT model (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) to perform a prediction task that classifies whether a review is fit-related and, consequently, more likely to have resulted from a consumer-product mismatch. Applying our classifier to the entire sample of reviews, we find that after monetization, the proportion of negative reviews directly related to product fit increases by 15% relative to baseline, indicating a worsened consumer-product mismatch. Furthermore, we present evidence that the magnitude of rating dispersion rises following the policy's introduction, providing additional support for the mismatch hypothesis.

In summary, our study's findings present a more intricate and nuanced perspective on platform monetization. Traditional wisdom regarding monetization often focuses solely on network effects, i.e., how the value of platform services depends on the number of participants on each side of the market. In our empirical context, the concern of network effects is that if a fee is introduced and the supply side (i.e., authors and publishers) of the market begins to dwindle, the other side (i.e., readers) may also cease to derive value, causing the market to unravel. We expand this view by examining the distribution and composition of different participants and products in this marketplace (authors, publishers, books, and readers).

By looking beyond the quantitative aspect of network effects, we identify some novel and (potentially) unintended consequences of monetizing the Giveaways program. These findings offer insights into the subtle yet crucial considerations that should be taken into account when evaluating the success of a monetization policy and the welfare of the ecosystem. Furthermore, they suggest that a platform must counterbalance undesirable effects by implementing more flexible incentive structures for the various players in its ecosystem.

2 Related Literature

Our paper resides at the nexus of several research domains, encompassing digitization, platform governance, product diversity, and consumer ratings. In this section, we synthesize the prevailing literature and underscore our contributions within these contexts

2.1 Digitization of Cultural Products

Digitization and the emergence of online platforms have created "a golden age of music, movies, books, and television" (Waldfogel, 2017), presenting both challenges and opportunities for creators. As entry barriers continue to decrease, we observe heightened cultural participation and an influx of new products within these markets. Focusing specifically on the book industry, we utilize it as a representative case to outline the primary effects of digitization on book market dynamics

On the supply side, Peukert and Reimers (2022) demonstrate that digitization yields valuable information for predicting book publishers' product success—specifically, license payment size more accurately reflects a book's ex-post success. Additionally, there is an evolving body of literature investigating the potential cannibalization of physical book demand due to digital distributionrevealing context-dependent outcomes. Exploiting an exogenous delay in the release of new Kindle e-books from a publisher, Chen et al. (2019) discover that postponing e-book availability does not lead to increased print book sales. Similarly, Nagaraj and Reimers (2021) finds that Google Books' digitization efforts actually enhance physical book sales by improving discoverability of existing works. In contrast, Sharma et al. (2021) estimates that digitization reduces print sales of adult fiction books—the e-book format's most popular genre—by up to 30% on average.

Although the supply-side dynamics of book digitization on physical books have received substantial research attention, there has been comparatively less focus on examining the effects of digital promotional strategies within the book market. As the array of new offerings available to consumers continues to expand, generating significant welfare benefits (Waldfogel, 2017; Aguiar and Waldfogel, 2018), self-promotion becomes increasingly crucial for content creators (authors and publishers in our context). Our study contributes to this literature by illustrating how the benefits derived from digital book promotions may be unevenly distributed. Specifically, we demonstrate how monetizing digital promotional tools can result in downstream consequences for the types of products being promoted

2.2 Platform Governance and Monetization

In numerous cultural product markets, digital platforms now serve as the primary mediators, resulting in a complex interdependence among various participants within the ecosystem. Specifically, platform decisions regarding which products to promote significantly impact the consumer product discovery process. Given the increasing importance of self-promotion for cultural products amid the crowded marketplace, platforms now wield considerable influence over the success of these products.

The platform governance literature is grounded in the concept that platforms manage a micro-economy and are therefore responsible for devising policies and mechanisms that foster desirable outcomes for their ecosystems (Boudreau and Hagiu, 2009; Tiwana et al., 2010; Parker et al., 2017; Parker and Van Alstyne, 2018). Prior research has explored various governance decisions, such as pricing mechanisms (Parker and Van Alstyne, 2005; Hagiu, 2006; Farronato, 2017), control and openness (Parker and Van Alstyne, 2018), platform awards and promotions (Rietveld et al., 2019; Foerderer et al., 2021), information exchange (Foerderer, 2020; Huang et al., 2019), and platform owner entry (Foerderer et al., 2018; Zhu and Liu, 2018). In the context of platforms' ability to influence multi-sided markets, Aguiar and Waldfogel (2021), for instance, investigate Spotify's promotion decisions on song and artist success, demonstrating that platforms hold considerable power over their ecosystems. Rietveld et al. (2019) argue that platforms strategically choose which complements to promote in order to effectively manage value creation within the entire ecosystem.

Among the numerous governance decisions a platform must make to manage value creation, developing a monetization strategy is crucial (Parker et al., 2016). Establishing a revenue stream is essential for the sustainable operation of a platform company. However, despite the complexity and multi-dimensionality of monetization, existing narratives primarily focus on preventing harm to network effects. Although network effects are important for growing two-sided markets, the overall health of platform ecosystems depends on multiple factors and may also influence the content produced within a creative market. In this regard, Wu and Zhu (2022) investigate how different revenue models in a novel-writing platform market respond to intensified competition. They discover that revenue-sharing books exhibit a greater response to competition than pay-by-the-word books in terms of content quantity and novelty. Our study expands this literature by offering new insights into the consequences of monetization beyond network effects concerns—we empirically examine outcomes such as supply concentration, product diversity, and consumer-product mismatch.

2.3 Product Diversity and Fair Representation

Our research investigates supply concentration and the subsequent decline in product diversity concerning book genres. Diversity has long been a vital scientific concept as well as a significant societal focus (Gini, 1921; Shannon, 1948; Cowell, 2000). Within the context of digital platforms, prior studies have shown that individual-level consumption diversity is strongly correlated with long-term user metrics (Waller and Anderson, 2019; Anderson et al., 2020). This raises a critical question in platform governance: how to guide user behavior to promote diverse consumption patterns while maintaining engagement (Anderson et al., 2020; Hansen et al., 2021).

A considerable amount of research has focused on understanding the impact of algorithmic recommendations on consumption diversity (Anderson et al., 2020; Holtz et al., 2020; Hansen et al., 2021). This body of work has found that recommendation systems often boost user engagement at the expense of decreasing consumption diversity, which may be detrimental to long-term platform success. Despite this concern, our understanding of how product and consumption diversity affect platform ecosystems, and the mechanisms through which diversity influences platform markets, remains limited.

Consumption diversity is closely related to the concept of product variety. Previous research has demonstrated that increased book variety, due to both online marketplaces (Brynjolfsson et al., 2003) and digital publishing and distribution (Waldfogel and Reimers, 2015), has created substantial welfare benefits for consumers. The phenomenon of increased product variety is often referred to as the long tail effect of the internet. In the context of books, Brynjolfsson et al. (2003) found that enhanced product variety in online bookstores increased consumer welfare by $731 million to $1.03 billion in the year 2000 alone. Given these figures, we anticipate that diverse offerings will play a crucial role in retaining and attracting consumers to Goodreads, a platform fundamentally reliant on user-generated content.

Moreover, diversity in production is closely linked to diversity and fair representation among creators. Low entry barriers for newcomers amplify supply-side competition, necessitating increased self-promotion by cultural producers who may rely on tools such as Giveaways to promote their work. Some scholars argue that the digital age embodies a revolutionary moment poised to democratize cultural production, disrupt patterns of inequality, and support underrepresented minorities (e.g., Toop (1995)). In the book market, for example, this could lead to greater participation by independent authors and smaller publishing houses. However, other scholars contend that the use of digital technology in cultural production will perpetuate existing workplace norms and biases (Grazian, 2005; Nagaraj and Ranganathan, 2022). Specifically, Nagaraj and Ranganathan (2022) find that digital recording technologies increase gender inequality for studio singers in the Indian Hindi film industry ("Bollywood"). They argue that this heightened inequality stems from differences in the extent to which men and women engage in self-promotion activities, which have become increasingly important in a crowded labor market for cultural products.

In our specific context, imposing a monetary cost for book promotion may disproportionately impact authors and publishers with limited resources, further intersecting with factors such as gender. To evaluate these opposing perspectives, our paper examines the diversity of offerings and the characteristics of authors and publishers to explore how monetizing a digital promotion tool like Giveaways affects the representation of creators and their work.

2.4 Price Discount and Consumer Ratings

Finally, our work directly examines the impact of promotional schemes, such as Giveaways, on subsequent product ratings. Previous research has shown an inverse relationship between promotions and ratings, with the influence of price discounts on consumer ratings garnering significant attention in the context of daily deal marketplaces (Byers et al., 2012a,b). Notably, Byers et al. (2012a) document a sharp decline in merchants' reputations on Yelp before and after running a daily deal. Moreover, Byers et al. (2012b) explore various hypotheses underlying these unintended consequences of deals. This phenomenon, now widely recognized as the Groupon effect, has become a stylized fact in the literature.

Recent research in book markets (Zegners, 2019) and mobile app markets (Liu et al., 2019) supports the negative impact of price discounts on consumer ratings—price discounts increase sales but result in lower consumer ratings for books and apps. This negative effect is particularly pronounced for free products (Liu et al., 2019). Interestingly, both studies suggest that the underlying mechanism for this phenomenon is a mismatch between consumer preferences and product characteristics: consumers who self-select to obtain products at a discounted price tend to have unfavorable idiosyncratic tastes for them, leading to a higher likelihood of providing lower ratings. Our paper also investigates the impact of price discounts on consumer ratings in our demand-side analysis but with a different emphasis. We aim to understand how the impact of monetization (through charging a fixed fee) on the supply side propagates to consumers and moderates the relationship between price discounts and consumer ratings in a platform market.

3 Setting and Natural Experiment

3.1 Empirical Setting

Our empirical setting is Goodreads, the world's largest online community for book readers. The platform allows users to search an extensive database of books, receive suggestions for future reading choices, curate their reading lists, and write book reviews. Users can also rate books on a scale of one to five stars, with the option to include a written text review alongside the rating. Goodreads is renowned for its vast repository of user-generated book reviews and star ratings, to the extent that its book ratings serve as the primary information source for book-related Google searches. Moreover, the platform incorporates social features, enabling users to connect with others and receive updates about their "friends" reading activities and book reviews. As the number of new titles increases annually, Goodreads plays an increasingly vital role in book discovery by assisting digital-age consumers in finding their next read and providing pertinent pre-purchase information, such as reviews and ratings.

In this paper, we concentrate on a platform-mediated feature within Goodreads that facilitates book promotion and discovery, specifically, the Giveaways program. Figure 1a illustrates the starting page for authors and publishers, while Figure 1b offers an example of a Giveaways campaign. Authors or publishers can initiate these campaigns, in which they pledge to distribute a specified number of their books (e.g., 100 copies in Figure 1b). Users can peruse individual Giveaways campaigns and choose to enter for an opportunity to win a free copy of the book. The winners are selected through a random drawing at the conclusion of the listing period (typically lasting a few weeks). Authors or publishers are then responsible for sending copies to the winners. The campaign also integrates social features—for instance, when a user's friend enters a campaign, it appears on their timeline, thereby increasing the book's visibility. Furthermore, this feature allows authors and publishers to distribute their books to non-professional reviewers (i.e., regular readers), who are then encouraged to provide a book review. The program is renowned for its diverse reading selections. Similar to browsing a book fair, users can discover promoted books spanning various genres, such as science fiction, romance, and biography. Over time, Giveaways campaigns have gained a reputation as exceptional tools for authors to generate buzz for their books, subsequently becoming a central marketing feature of the Goodreads platform.[2]

Previously, the Giveaways program was free for authors and publishers to participate in. However, this changed in January 2018 when Goodreads began monetizing the program by charging a flat participation fee to its supply side.[3] The fixed price is $119 for a single Giveaways campaign,[4] with the cost doubling for promoting two books. The policy announcement was made on November 28, 2017, and took effect on January 9, 2018. Authors and publishers did not anticipate this change,[5] and the short time span between the policy announcement and its implementation did not allow them to react - for instance, it is unlikely that participants strategically shifted theirGiveaways to the pre-monetization period.[6] The exogenous vari-

[2]For more information, see https://www.goodreads.com/giveaway

[3]For more information, see the archived Goodreads blog announcement: https://web.archive.org/web/20171215071901/https:/www.goodreads.com/blog/show/1108-goodreads-introduces-new-u-s-giveaways-program-a-more-powerful-book-mar.

[4]For the premium version, which features the paying author's or publisher's Giveaways campaign on the first page, the cost is $599.

[5]This is further supported by tweets, Reddit posts, and news articles where authors discuss their surprise and reactions to this change; see https://www.reddit.com/r/Fantasy/comments/7yym6y/authors_opinions_on_new_goodreads_giveaway_fees/.

[6]This is further evident from the lack of bunching in Figure 2—we do not see a sharp increase in Giveaways participation in the months preceding January 2018. Nevertheless, for our supply-side analysis, we exclude a three month window around monetization, i.e., January 1 to March 31, 2018, to allow for a longer adjustment period. Additionally, we only consider a Giveaways campaign if the format of the giveaway books is a printed edition.

ation induced by this natural experiment offers a clear setting in which we can investigate how platform monetization impacts the supply side, product diversity, and, ultimately, consumer satisfaction.

Figure 1: Illustrating the Giveaways Program

(a) Starting Page of the Giveaways Program for Authors and Publishers

(b) Example of a Giveaways campaign

Notes: The figure presents screenshots of the starting page and an example of a Giveaways campaign featuring the book title "Greenwich Park." These images highlight some key features of the program, offering a visual representation of how users can engage with the content and participate in Giveaways campaigns.

3.2 Data

Employing both the Goodreads Application Programming Interface (API) and a customized web crawler, we gather comprehensive and in-depth data from various sources to examine the consequences of this exogenous monetization policy shift. Specifically, we compile: (1) the entire set of Giveaways campaigns hosted on the platform from January 2016 to February 2020, prior to the pandemic—approximately two years before and two years after the implementation of the monetization policy in January 2018; (2) metadata for all books involved in those Giveaways campaigns, in addition to information about their authors and publishers; and

(3) star ratings and textual reviews associated with each book promoted through the Giveaways program.

In constructing our estimation sample, we employ two sample-filtering criteria on the raw data. First, we exclude books published more than five years prior to their participation in Giveaways and those not published within one year after participation. The average difference between a book's participation date in Giveaways and its release date is 108 days, indicating that a typical participating book is promoted through the Giveaways program 3-4 months post-release. We eliminate extreme cases, as the decision to promote these books may significantly differ from promoting a book 3-4 months after its release. This step removes approximately 2.5% of books, leaving 87,966 books in the sample. Second, we exclude books that participated in Giveaways more than three times, as the effects of program participation may greatly vary for these books compared to those participating only once or twice. This eliminates an additional 6% of books.[7]

Table 1 presents a summary of our sample after applying the two exclusions, comprising 101,684 giveaway events for 82,552 books, with nearly 54 million ratings associated with those books.[8] Through these Giveaways campaigns, 1,814,643 free book copies have been distributed to readers. The basic summary statistics for book-/author-level metadata and ratings data are detailed in Table 2. Our analysis will primarily focus on Giveaways involving printed books, which constitute the vast majority (over 94%) of hosted Giveaways (results remain qualitatively similar when e-books are included).

Table 1: Summary of the Giveaways Sample

	Total Number
Books	82,552
Giveaway Campaigns	101,684
Free Copies of Books	1,814,643
Unique Authors	66,785
Unique Publishers	15,323

Table 2: Descriptive Statistics for Rating and Review of Books and Authors

	Mean	Median	Std	Min	Max
Books					
Average Rating	4.025	4.030	0.449	1	5
Number of Ratings	665.1	22	6,711.512	0	780,109
Number of Reviews	91.39	8	668.1431	0	57,042
Authors					
Average Ratings	4.01	4.0	0.361	1	5
Number of ratings	51,791	516	381,951.5	0	25,840,312
Number of Reviews	3,640	116	19,331.08	0	588,564

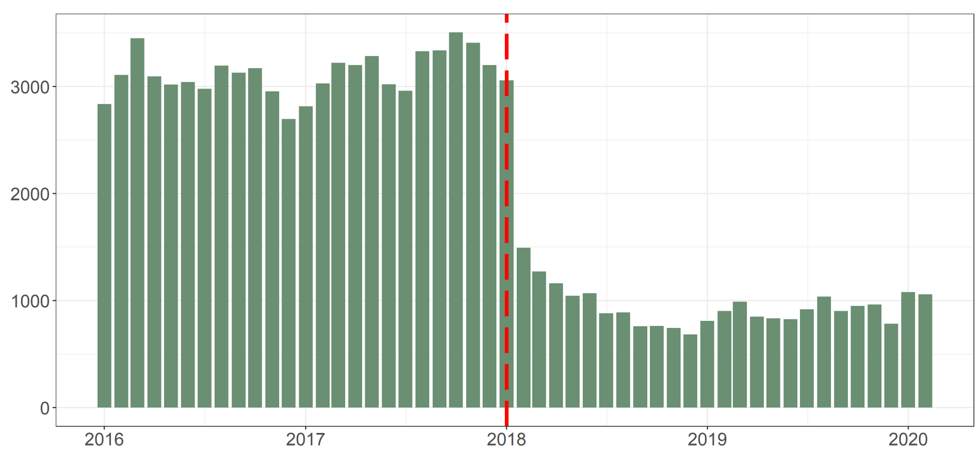
[7] Our results remain robust across different data subsets.

[8] One limitation of our rating data is that Goodreads only permits scraping a maximum of 3,000 individual reviews and ratings per book. However, when comparing our collected sample to the raw numbers of 1-5 star ratings obtained through book-level metadata, we find highly similar distributions, thus providing evidence that our sample accurately represents books' review distributions. This distribution is available in Appendix B.

3.3 Model-Free Evidence

A priori, without examining the data, it remains uncertain whether a few hundred dollars would present a significant obstacle for authors and publishers, considering the effort required to create a new book. While major publishing houses like Penguin Random House boast of extensive resources, independent authors collaborating with smaller publishers or utilizing self-publishing platforms often have restricted budgets. As a result, the participation fees imposed by the platform might emerge as a substantial barrier for these authors seeking to engage in this program. Indeed, our preliminary empirical observation indicates that the monetization of the Giveaways program in January 2018 considerably affected participation from authors and publishers. In Figure 2, we observe an immediate and substantial decline in the monthly number of campaigns on the platform upon the introduction of the new monetization policy, with the average number of Giveaways campaigns per month plummeting from around 3,000 to a mere 1,000. This initial evidence suggests that the monetary burden poses a significant obstacle, especially for certain authors and publishers, and therefore carries implications for book supply.

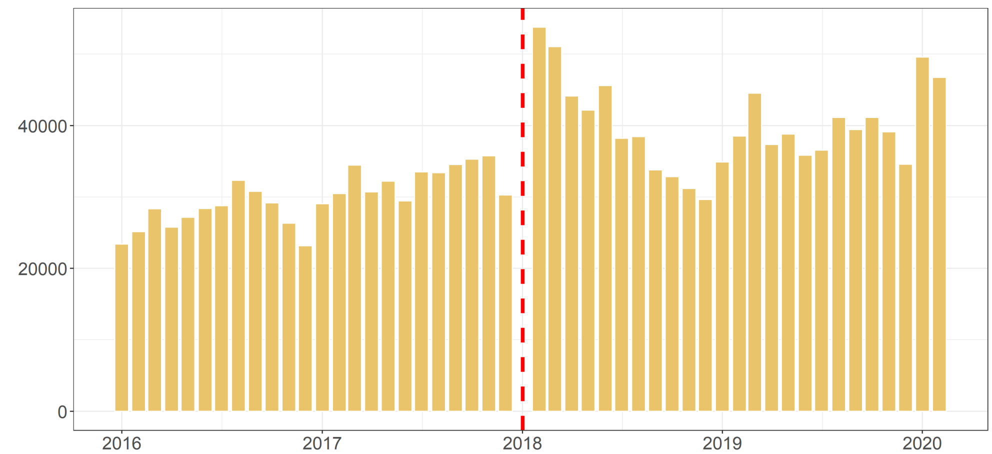
Figure 2: Giveaways Campaigns Over Time: The Impact of Giveaways Monetization



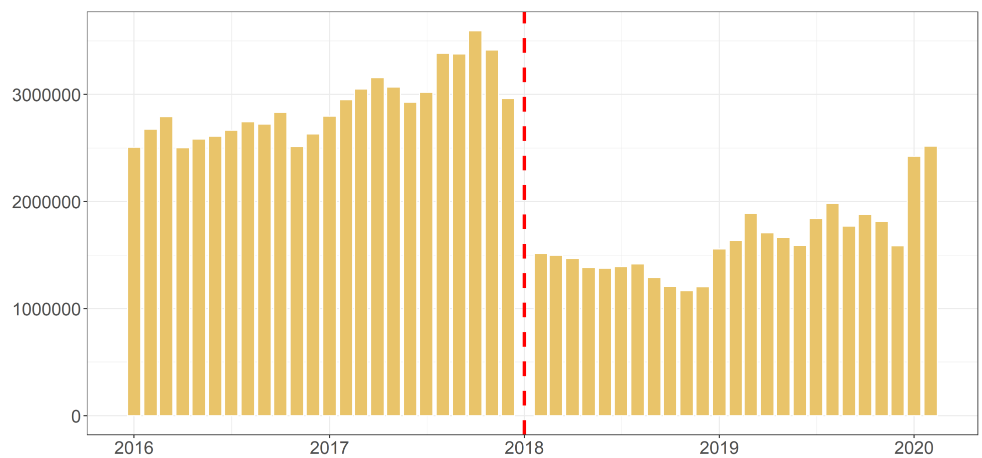
Notes: A sharp decline in participation is observed following the implemention of monetization. The red vertical dashed line marks January 2018, when the platform began to impose fees for participation in Giveaways campaigns.

Filtering out some participation might not necessarily be a negative outcome. We observe in our data that there are some "bad actors" who repetitively post their books in the Giveaways program many times over a certain period, giving away only a few copies each time. These authors are likely trying to gain exposure by consistently appearing on the listing page of the Giveaways program, and a participation fee would deter such actors from exploiting the system in this manner. However, based on the data we collect, authors or publishers engaging in repetitive posting more than three times account for only 6% of participating books, indicating that such gaming behavior is relatively rare. In any case, a decrease in supply at this scale is still cause for concern. Later in this paper, we will investigate the sources of this reduced supply participation and the implications it has for this book promotion marketplace. Furthermore, we will examine whether the quality of books (as measured by star ratings) significantly differs for books participating prevs. post-monetization, and we find no evidence to support this.[9]

Figure 3: Book Copies Offered and Readers' Requests in Giveaways: Impact of Giveaways Monetization



(a) Supply Side: Number of Book Copies Provided by Publishers and Authors.



(b) Demand Side: Number of Readers' Requests for Giveaway Books.

Notes: Panel (b) shows a nearly 50% decline in readers' engagement with the Giveaways program, as indicated by the number of requests submitted for giveaway books, following January 2018. This decrease occurs despite an increase in the overall number of book copies available within the Giveaways ecosystem, as shown in Panel (a). The surge in available copies is driven by authors being motivated to offer more copies of their books per campaign when confronted with a fixed participation fee. The red vertical dashed line marks January 2018.

We next examine whether readers alter their behavior in response to the monetization policy change. As depicted in panel (b) of Figure 3, readers' participation in the Giveaways program diminishes by almost half: the number of book requests from readers declines from around 3 million per month to approximately 1.5

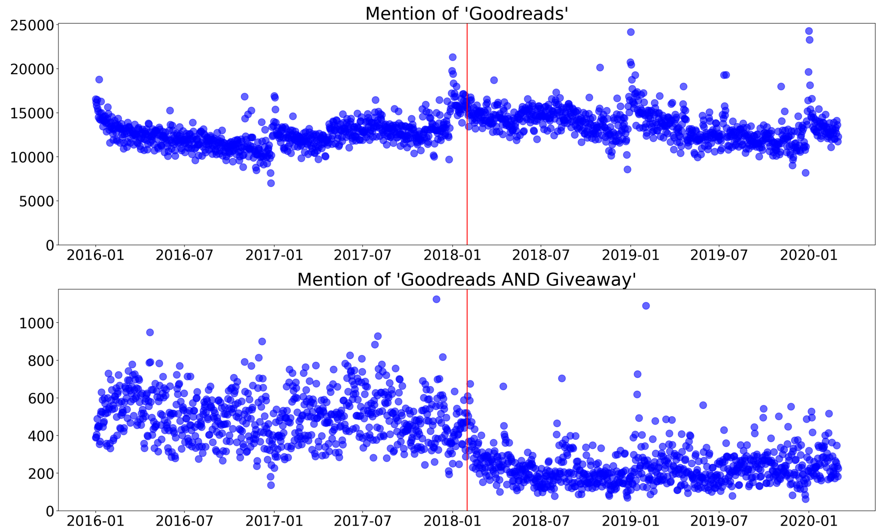
[9] It is worth noting that we do not find any evidence to suggest that the quality of books, as measured by star ratings, differs significantly between those participating pre- vs. post-monetization.

million.[10] This reduction occurs despite an increase in overall book copies available within the Giveaways ecosystem (as shown in panel (a) of Figure 3), which is driven by authors being incentivized to provide more copies of their book per campaign when faced with a fixed participation fee. This result illustrates the complexity of monetizing a two-sided market: strategies that directly impact one side may also have spillover effects on the other side of the market. Considering that a primary objective of the Giveaways program is to build an audience for the promoted books and encourage readers to discuss them, the decrease in overall participation from readers may negatively affect word of mouth, which is undesirable from the platform's perspective.

In addition to generating buzz and engagement on the Goodreads platform, word-of-mouth promotion on social media platforms, such as Twitter, can also raise awareness and visibility for featured books (Adamopoulos et al., 2018). To evaluate the effects of monetization on word-of-mouth promotion beyond Goodreads, we present some aggregated evidence from Twitter data. The lower panel of Figure 4 demonstrates a marked decline in the volume of Twitter discussions mentioning both "Goodreads" and "Giveaways" around the time the monetization policy is implemented. Along with the overall volume of Twitter conversations, engagement in the form of comments and retweets also declines (see Appendix A). It is crucial to recognize that this decrease in word-of-mouth promotion is not attributable to any platform-wide effects, as the volume of tweets discussing "Goodreads" remains relatively stable during this period, as illustrated in the upper panel of Figure 4. This suggests that the impact is specific to the Giveaways program. Our analysis of word-of-mouth promotion and engagement on Twitter, in conjunction with the observed decline in program participation, offers preliminary insight into the potential drawbacks of monetizing Giveaways for the platform. The remainder of this paper will further explore the implications for the platform marketplace.

[10] In Figure 3, we exclude both the number of book copies available on Giveaways (97,777) and the number of readers' requests (2,868,526) for January 2018. The substantial quantity of free book copies in January 2018 is a result of the rising count of giveaway book copies per campaign, while the number of campaigns hosted during that month remained relatively constant.

Figure 4: Daily Tweet Count Reflecting Word of Mouth: The Impact of Giveaways Monetization



Notes: The figure illustrates the effect of introducing a flat participation fee on the word-of-month of the Giveaways program beyond Goodreads. As shown in the lower panel, there is a decrease in the number of Twitter discussions mentioning both "Goodreads" and "Giveaways" following the implementation of the monetization policy, while the number of tweets discussing "Goodreads" remains relatively stable throughout the entire period as demonstrated in the upper panel. The red vertical line marks January 2018.

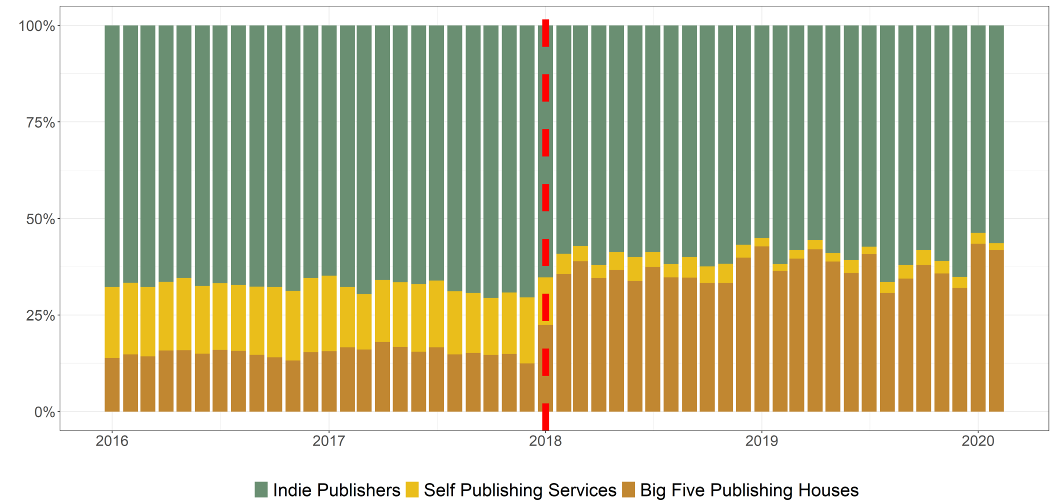
4 Empirical Results

In this section, we explore the implications of monetizing the Giveaways program on the platform ecosystem by analyzing the impact of a flat participation fee on three key aspects: (i) supply-side effects, examining how the fee impacts authors and publishers running promotional campaigns for books within the program; (ii) product availability, assessing the ways in which monetization shapes the variety of books offered in the program; and (iii) demand-side effects, investigating how readers react to supply-side alterations prompted by changes in (i) and (ii).

4.1 Supply Side Dynamics: Effects on Publishers and Authors

Impact on Publishers. In order to contextualize our supply-side analysis, we begin by offering an overview of the publishing industry. Traditionally, a book must first be accepted by a publishing house editor before the publisher acquires the exclusive publishing rights from the author and proceeds with publication. Within this realm of publishing services, the Big Five publishers—Penguin Random House, Hachette Livre, HarperCollins, Simon & Schuster, and Macmillan Publishers—hold significant market dominance, accounting for over 60% of all English-language titles published in the US. Apart from the Big Five, smaller "indie" publishers also adhere to the traditional publishing model. Moreover, advances in digital technology have ushered in self-publishing, a novel option that enables authors to invest their own resources in publishing their books, both online and offline. Each publisher category has its own unique publishing processes and utilizes distinct promotional channels for their publications. Consequently, the monetization of the Giveaways program may yield varying implications for each group.

Figure 5: Publisher Categories on Giveaways Over Time: The Impact of Giveaways Monetization



Notes: Percentage of Giveaways Campaigns Participated by Different Publisher Categories Over Time. Although all publishers decreased their number of campaigns on Giveaways, this figure demonstrates an increase in the percentage of books by Big Five publishing houses and a decline in the percentages of books by indie publishers and self-publishing services. The red vertical dashed line marks January 2018.

We collect publishing information for each book featured in the Giveaways campaigns and associate each book with a specific publisher or publishing service.[11] To understand the impact of imposing a participation fee on the Giveaways program for various publisher types, we divide them into three categories based on the discussion above: (i) Big Five publishers, (ii) indie publishers, and (iii) self-publishing service providers.

Following the introduction of participation fees on Giveaways, all publisher types saw a marked decline in the number of monthly campaigns. Prior to monetization, Big Five publishing houses hosted about 469 campaigns per month, which dropped to 249 afterward. Self-publishing services, which initially hosted more campaigns than Big Five publishing houses with 529 monthly campaigns, experienced a sharp decrease to only 22 campaigns per month following the policy change. Small indie publishers' campaigns went from 2,083 books per month to a mere 400. The impact of the monetization policy varies among different publisher types. Figure 5 provides a more detailed view of the effect on publisher composition. Following monetization, the proportions of books by indie publishers and self-publishing services decrease, while the proportion of books from the Big Five publishing houses more than doubles (from 12% to 30%). This outcome highlights the varying effects that the monetization of Giveaways has on different publishers. To further examine this heterogeneity, we aggregate Giveaways books for each publisher and perform an analysis at the publisher level. We estimate a regression model using the following form:

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 \times \text{Post-Monetization}_t + \beta_2 \times \text{Post-Monetization}_t \times \text{Big-Five}_i \\ + \beta_3 \times \text{Post-Monetization}_t \times \text{Self-publishing}_t + \delta_t + \sigma_i + \epsilon_{it}, \end{aligned} \tag{1}$$

where Post-Monetization$_t$ is a dummy variable equal to 1 if time t is after Goodreads modified its

[11]More details on these classifications and our data can be found in Section A2.

Giveaways participation policy in January 2018, Big-Five$_i$ indicates whether publisher i is among the Big Five publishing houses, and Self-publishing$_i$ denotes if a book is self-published. We estimate the above model using two dependent variables y_{it}: the number of Giveaways campaigns publisher i has in month t and the proportion of books in the Giveaways program by publisher i in month t, denoted as Prop$_{it}$.[12] δ_t represents month fixed effects, and σ_i represents publisher fixed effects. β_1 illustrates the impact of Giveaways monetization relative to the baseline group—indie publishers. Given that the observations of one publisher can be correlated, we cluster the standard errors at the publisher level.

The result in Table 3 align with the findings from Figure 5. Column (1) reveals that all publishers reduce their participation in the Giveaways program following its monetization: indie publishers' participation decreases by 0.618 campaigns each month ($\beta = -0.618, p < 0.01$), Big Five publishers' by over 42 ($\beta = -42.8, p < 0.01$), self-publishing providers' by 53 (although $p > 0.1$). It should be noted that there are only a few popular self-publishing channels, resulting in limited statistical power for this category - we visually show in Section C.2 that the decrease in self-publishers' participation is greater even in absolute terms. In terms of publisher proportions in the monthly compositions, however, the proportion of books by Big Five publishing houses increases by 0.5% ($\beta = 0.005, p < 0.01$) post-monetization. This outcome suggests that self-publishing service providers and indie publishers are disproportionately affected by the participation fee compared to Big Five publishing houses. Authors associated with smaller publishers or self-publishing services, who possess fewer marketing tools for their books, are most significantly impacted by the policy change due to their limited resources. [13]

Changes in Supply Concentration. The aforementioned results on publishers imply that the Big Five are expanding their share in the Giveaways marketplace at the expense of smaller publishers and self-publishing service providers, consequently increasing market concentration. We employ the Herfindahl-Hirschman Index (HHI) as a measure to quantify the change in publisher concentration over time. We calculate the HHI for publishers in each month of our data, where HHI_t is given by $\sum_{i=1}^{n} s_{it}^2$ and s_{it} denotes the market share of publisher i in month t. A larger HHI signifies greater publisher concentration (Hirshman, 1964). Figure 6 demonstrates that the HHI indeed rises after monetization by 2.5 times (from 0.01 to 0.035), indicating a concentration of power among a few publishers. In other words, monetization not only results in a decrease in the overall participation of books in Giveaways but also affects the mix of participating publisher types.

Impact on Authors. We now present evidence on how Giveaways monetization impacts individual authors by examining the characteristics of authors participating in the Giveaways program before and after January 2018. Panels (a)-(d) of Figure 7 reveal that post-monetization, authors who continue participating in Giveaways tend to be more popular, experienced, and recognized than those who participated prior to the change. On average, they garner more ratings and reviews for their books and maintain larger networks (i.e., more friends) on Goodreads. Moreover, these authors demonstrate greater familiarity with the Giveaways program, as evidenced by their increased participation in campaigns during the same timeframe.

[12] We first calculate Percentage$_{it} = \frac{N_{it}}{\sum_j N_{jt}}$, where j represents all publishers including publisher i. Then, as a standardization, we divide Percentage$_{it}$ by the mean of Percentage$_{it}$ before the monetization policy change, and thus Prop$_{it} = \frac{Percentage_{it}}{\frac{\sum_{t'} Percentage_{it'}}{T}}$, where t' indicates the time period before Giveaways started charges for participation and T is the number of periods pre-monetization.

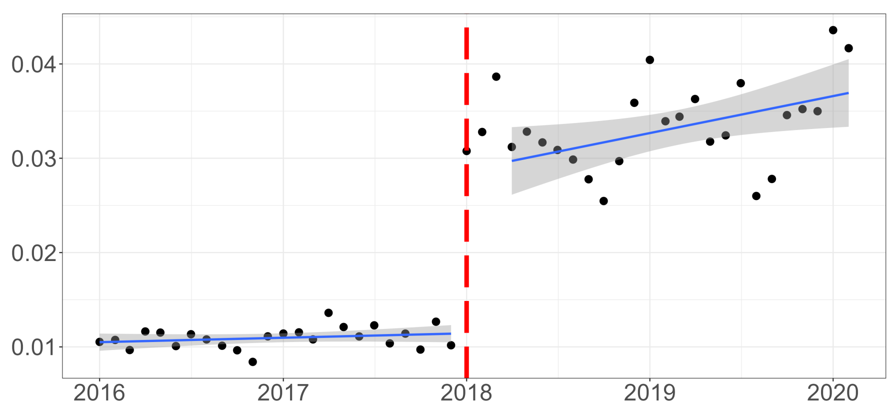
[13] To further corroborate the results regarding how the mix of books on the market changes in terms of publisher characteristics pre- versus post-monetization, we conduct a book-level analysis and estimate how the likelihood of Giveaways participation differs for a book published by a Big Five publisher vs self-publishing service after monetization. Section C.3 demonstrates that the likelihood of a Big 5 book being put on Giveaways increases by 11%, while that of a self-published book decreases by 5% following monetization.

Table 3: Effects of Giveaway Monetization on Book Supply across Publisher Categories

	Dependent Variable:	
	Number of Campaigns	Proportion of Campaigns
	(1)	(2)
Post-Monetization	-0.618^{***}	-0.025^{***}
	(0.034)	(0.001)
Post-Monetization × Big 5	-42.825^{***}	0.005^{***}
	(6.267)	(0.002)
Post-Monetization × Self-Publisher	-53.930	-0.008^{**}
	(42.362)	(0.004)
Month Fixed Effects	Yes	Yes
Publisher Fixed Effects	Yes	Yes
Observations	79,242	79,242
R^2	0.727	0.092

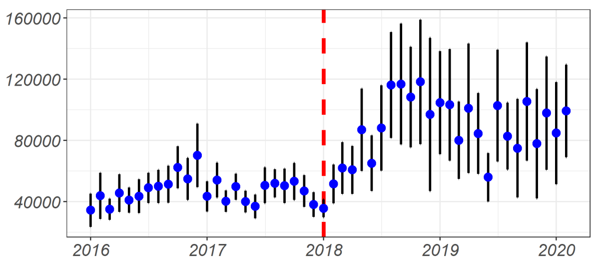
Note: *p<0.1; **p<0.05; ***p<0.01

Figure 6: Publisher Concentration on Giveaways Over Time: The Impact of Giveaways Monetization

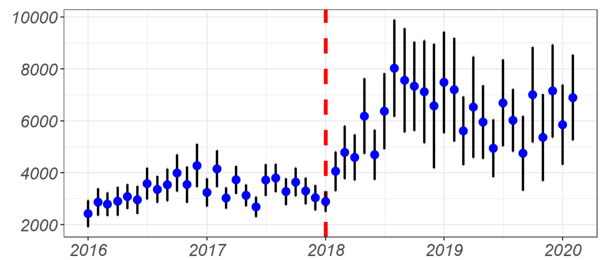


Notes: HHI to Depict the Change of Publisher Concentration Over Time. There is a sharp rise in HHI, signifying that publisher concentration experiences a significant increase following monetization. The red vertical dashed line marks January 2018, and the grey band indicates the 95% confidence interval.

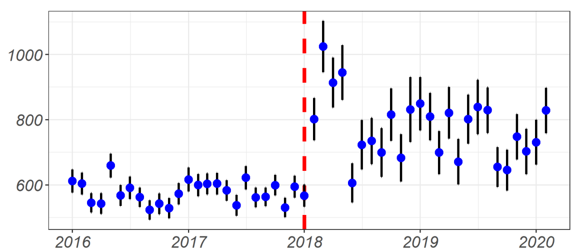
Figure 7: Author Characteristics on Giveaways Over Time: The Impact of Giveaways Monetization



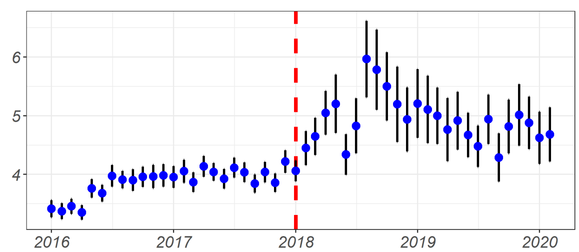
(a) Number of Ratings Authors Received



(b) Number of Reviews Authors Received



(c) Number of Friends on Goodreads



(d) Prior Number of Giveaways Authors Participated

Notes: The four graphs display the average monthly ratings and reviews received by authors, the average number of Giveaways campaigns they participate in, and the number of friends they have on Goodreads. The red vertical dashed line identifies January 2018, and the black bars indicate the 95% confidence interval.

This evidence complements our publisher-level analysis, suggesting that established authors (with more reviews and a prominent Goodreads presence) are more likely to continue using the Giveaways program after monetization, while their less experienced counterparts with weaker track records may discontinue participation.[14]

To investigate whether the policy has a disparate impact across author genders, we assess the relative changes in participation by male and female authors. First, we predict the gender of authors utilizing U.S. Census data and U.S. Social Security Administration data.[15] Next, we tally the monthly number of books participating in the Giveaways program by male and female authors, and estimate a regression model analogous to the above publisher analysis.

$$\mathrm{y}_{kt} = \beta_0 + \beta_1 \times \text{Post-Monetization}_t + \beta_2 \times \text{Post-Monetization}_t \times \text{Female-author}_k + \delta_t + \epsilon_{kt}, \tag{2}$$

where y_{kt} denotes the number of Giveaways campaigns by author of gender k in month t, and Female-author$_k$ equals 1 for female authors. δ_t captures month fixed effects. As demonstrated in Table 4, the model estimates indicate that female authors are disproportionately affected by the new monetization policy. In comparison to male authors, female authors experience a reduction of 186.8 ($p < 0.01$) book campaigns per month following monetization. The table also reveals that overall, female authors utilize the Giveaways program more frequently to promote their books, initiating 297 ($p < 0.01$) additional book campaigns on average each month. These findings highlight the unequal impact of the flat fee on publishers and authors, with female authors—who rely more heavily on the program—bearing the brunt of the change.

Table 4: Effects of Giveaways Monetization on Book Supply across Author Gender

	Dependent Variable: No. of Book Campaigns
	(1)
Post	$-1{,}010.7^{***}$
	(24.1)
Female	297.0^{***}
	(23.9)
Post × Female	-186.8^{***}
	(34.1)
Month Fixed Effects	Yes
Observations	94
R^2	1.0

Note: *p<0.1; **p<0.05; ***p<0.01

In summary, our findings reveal that the monetization policy significantly influences both publishers' and authors' utilization of the Giveaways program, with varying effects across different groups. While

[14] It should be noted, however, that the quality of authors as measured by average ratings remains unchanged, as demonstrated in Figure 12.

[15] The R package Predictrace aids in this process. More information can be found at `https://cran.r-project.org/web/packages/predictrace/index.html`.

there is a substantial decline in participation from self-publishing service providers and indie publishers, the proportion of books by Big Five publishing houses actually rises following monetization. Moreover, female authors tend to use Giveaways more frequently for self-promotion compared to male authors; however, their books are more adversely affected by the monetization policy. These results underscore the need to consider the heterogeneous impacts of such policies on different segments of the publishing industry.

4.2 Exploring Product Diversity: An Analysis of Book Genres

In this section, we investigate the impact of the new participation fee on books enrolled in the Giveaways program, with a focus on product variety in terms of genre diversity. Diversity is particularly relevant for markets of cultural products, such as books, due to the horizontal differentiation arising from consumers' idiosyncratic tastes. Offering a diverse selection of books appeals to readers with diverse interests and addresses potential variety-seeking behavior. Moreover, greater diversity is associated with larger cross-side network effects for readers, as the supply side becomes collectively more attractive. As a result, product diversity, measured by book genre diversity, is an essential quality for the platform ecosystem and its users (the value of a long tail in online assortments is well established in the literature, e.g., Brynjolfsson et al., 2003).

To assess whether and how the monetization of Giveaways affects genre diversity, we first deduce genres for each book by examining the most frequent shelves they were added to on Goodreads.[16] Following this, we adopt Shannon entropy as our diversity metric for genre distribution. Stemming from Shannon's information theory (Shannon, 1948), entropy has been widely utilized as a diversity measure across various disciplines, also known as the Shannon diversity index. This quantitative measure reflects the number of distinct types present and the distribution of individuals among those types.

For each month, we compute the Shannon entropy value for the collection of books participating in the Giveaways program during that month. Figure 8 visualizes the mean and standard deviation of a monthly entropy-based diversity index for book genres before and after the monetization of the Giveaways program. We observe a clear and significant drop in genre diversity following January 2018. We also obtain similar results when controlling for potential temporal dynamics with a time trend term and month fixed effect estimation (reported in Appendix D). These results are also replicated using alternative concentration measures (e.g., Gini coefficient, HHI).

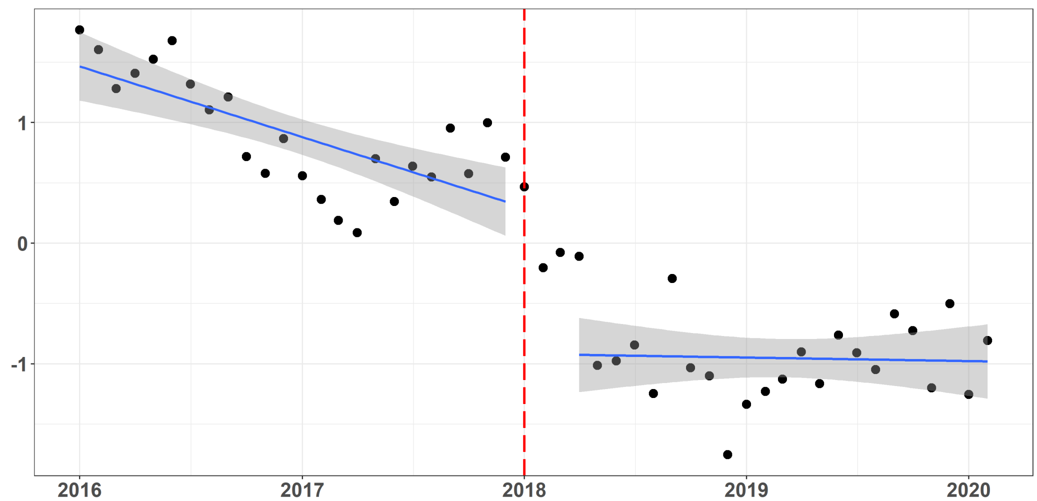
To investigate the mechanism driving the decrease in genre diversity following the program's monetization, we dive deeper into the impact on each specific book genre. Figure 9 displays a few illustrative examples of genre proportion changes before and after monetization, including three popular genres (first panel) and three niche genres (second panel). In the first row of the figure, we see thriller, mystery, and historical fiction - all genres with relatively high proportions among giveaway books prior to monetization, and their proportions increase further afterward. In the second row, the baseline proportions of science, psychology, and poetry are low before monetization and decline even lower after. The crucial observation here is that a few popular genres become more dominant in genre proportions following monetization, while the proportion of niche genres diminishes further shrinks.[17]

To rigorously test this observation, we also perform regression analysis on data from all 50 genres to scrutinize the heterogeneous impact of monetization across different genre types. To achieve this, we estimate the following model:

[16] Appendix D provides more details about this process.

[17] The complete set of proportion change plots for all 50 genres can be found in Appendix D, Figure 17.

Figure 8: Entropy-Based Genre Diversity Over Time: Impact of Giveaways Monetization



Notes: The analysis reveals a decrease in entropy during the post-monetization period, indicating reduced genre diversity following the implementation of the monetization policy. The red vertical dashed line identifies January 2018, and the grey band indicates the 95% confidence interval.

$$\begin{aligned}\text{Genre Proportion}_{gt} = \beta_1 \times \text{Post-Monetization}_{gt} \times \text{Top 25\% Genre}_g + \\ \beta_2 \times \text{Post-Monetization}_{gt} \times \text{Bottom 25\% Genre}_g + \gamma_m + \epsilon_{gt},\end{aligned} \tag{3}$$

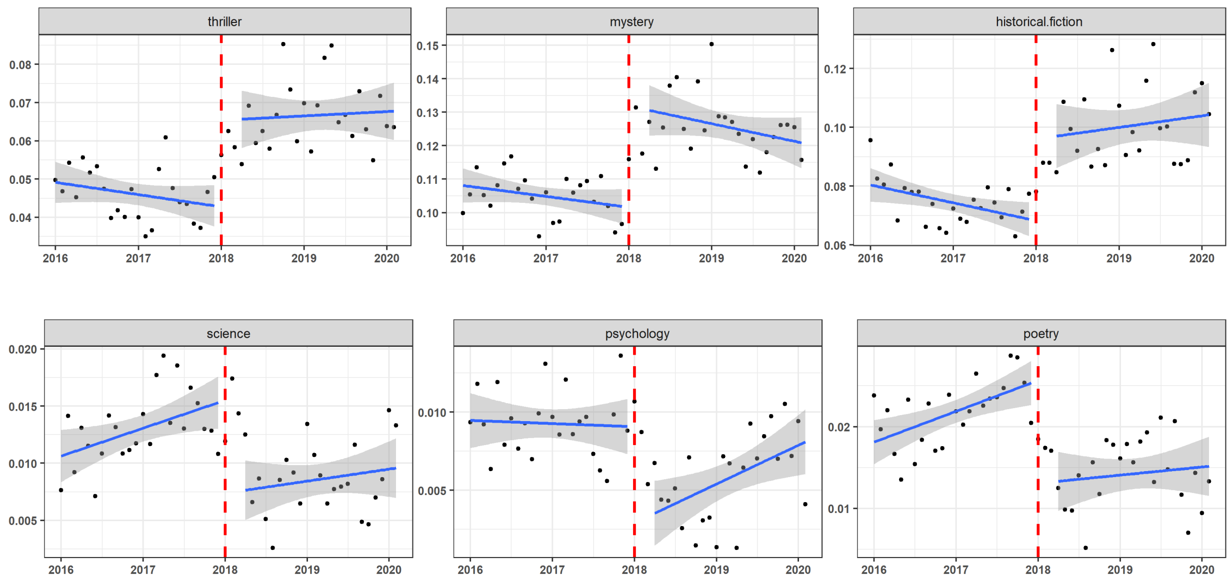
where the dependent variable is the genre proportion of genre g at giveaway month t. We categorize genres into three groups based on their quartiles in the distribution of genre proportions during the pre-monetization period and create binary indicators for them accordingly: top quartile (i.e., top 25% genre-popular group), bottom quartile (i.e., bottom 25% genre-niche group), and anything in between (i.e., quartile of the genre falls between 25% and 75%-middle group).

Table 5 presents the results: the model estimates are consistent with the visual examples in Figure 9 and support the bipolar effect on book genres. In Column (1), the model estimates reveal that post-monetization, the proportions of popular genres rise by 1.2% on average, while those of niche genres decline by 0.1% on average. Both estimates are statistically significant when compared to the middle group. Niche genres have a low baseline, so even a small change in absolute terms could have a substantial impact. To gauge the importance of change in proportion relative to each genre's own baseline, we also calculate the percentage change in proportion and employ it as an alternative dependent variable in the model.[18] Column (2) demostrates that the proportion of niche genres decreases by as much as 26%-a considerable decline relative to their own baseline. In contrast, a 1.2% increase in raw proportion corresponds to an 11%post-monetization increase for popular genres.

In summary, our findings provide evidence that monetizing the Giveaways program leads to a substantial

[18] For example, if romance has a genre proportion of 33% in November 2019 and its average pre-period proportion is 30%, the percentage change data point for November 2019 would be 33%/30% = 110%, indicating a 10% increase relative to its pre-period baseline.

Figure 9: Heterogeneous Impacts on Book Genres: Widening the Gap Between Popular and Niche Genres



Notes: We observe a growing disparity between popular and niche book genres as a result of the platform's monetization policy. The red vertical dashed line identifies January 2018, and the grey band indicates the 95% confidence interval.

Table 5: Effects of Giveaways Monetization on Popular and Niche Book Genre

	Dependent Variable:	
	Raw Proportion	Percentage Change
	(1)	(2)
Post-Monetization × Top 25% Genre	0.012^{***}	0.110^{***}
	(0.001)	(0.010)
Post-Monetization × Bottom 25% Genre	-0.001^{***}	-0.262^{***}
	(0.0001)	(0.039)
Month Fixed Effects	Yes	Yes
Genre Fixed Effects	Yes	Yes
Observations	2,350	2,350
R^2	0.969	0.188

Note: *p<0.1; **p<0.05; ***p<0.01

decline in the genre diversity of books offered through the program. This decrease can be attributed to the fact that genres already popular prior to monetization gain market share at the expense of less popular genres, which experience an even lower proportion following the introduction of a participation fee. Overall, these results emphasize that the policy fosters a rich-get-richer, poor-get-poorer dynamic in book genre diversity. This outcome may be a consequence of a selected group of publishers and authors being able to continue leveraging the Giveaways program for promoting their work post-monetization, as illustrated in the previous section discussing supply-side changes. Furthermore, even for those publishers and authors who continue using Giveaways after monetization, they may alter their behavior by promoting only mainstream, widely appealing books.

4.3 Demand Side of the Market: Consumer Ratings and Book Reviews

In this section, we investigate if the supply-side changes and book genre shifts resulting from the monetization of the Giveaways program extend to the demand side of the market, influencing consumer ratings and book reviews. Addressing this question offers insights into how these observed effects impact consumers, as well as providing valuable managerial implications for digital platforms.

Existing literature on promotional effects has established a general pattern in which price discounts typically lead to increased product adoption, but often at the expense of attracting lower online ratings (e.g., studies of the Groupon effect by Byers et al., 2012b). More recent research on this topic (Liu et al., 2019; Zegners, 2019) has explored preference mismatch as a mechanism driving this effect. In other words, consumers may not be the ideal fit or have the right taste for the discounted products, but they still opt to consume them due to the reduced cost. Consequently, they are more likely to be disappointed and leave lower ratings. In our study, we also examine this promotional effect in the context of Giveaways campaigns, which promote books at lower prices (i.e., free). However, our investigation's focus diverges from and extends beyond the promotional effect: our primary interest lies in understanding how the platform monetization policy on the supply side of the Giveaways program is transmitted to consumers and how it moderates the relationship between price discounts and consumer ratings. The summary statistics for the full set of individual reviews and ratings data collected from Goodreads is provided in Table 6.

Table 6: Summary of Scraped Ratings Data

Sample	Total Books	Counts of Ratings	Mean	Median	Std	Max	Min
Disaggregated Sample	81,608	36,631,238	3.47	4	1.20	5	1
Book Level Average	81,608	448.87	3.92	4.02	0.96	4.96	1.70

To examine the effect of Giveaways monetization on consumer ratings and reviews, we estimate the following fixed effect regression model at the book-month level:

$$r_{jt} = \alpha_j + \gamma_t + \beta_1 \times \text{Post-Giveaway}_{jt} + \beta_2 \times \text{Post-Giveaway}_{jt} \times \text{Post-Monetization}_j + \epsilon_{jt}, \tag{4}$$

where r_{jt} represents our dependent variable, denoting either the average rating score or the number of ratings for book j in month t. Post-Giveaway is an indicator variable equal to 1 if month t occurs after the end of the Giveaways campaign for book j, while Post-Monetization is an indicator variable equal to 1 for Giveaways campaigns occurring in month t after the implementation of the monetization policy in January 2018. We further include book fixed effects to account for time-invariant book characteristics and year-month fixed effects to account for common temporal trends in demand across books. Standard errors are clustered at the book level.

The parameter of interest, β_2, represents the impact of the Giveaways program monetization on the volume and valence of ratings for books promoted through this channel. In particular, β_2 estimates the difference in promotional effects for books participating in Giveaways post-monetization compared to those participating pre-monetization. This can be interpreted as a difference-in-difference estimator, where pre-monetization

Giveaways books serve as the control group and form the basis for counterfactual estimation. We examine the selection on observables assumption and perform additional trend analyses in Appendix E (Figure 18a and Figure 18b). In our primary analysis, we focus on a two-year window surrounding the Giveaways participation date (12 months before and 12 months after) to reduce the influence of non-Giveaways-induced ratings and long-term trend effects[19]. Our results, presented in Appendix E (Table 13 and Table 14), remain consistent when all data points are considered.[20]

Our primary finding reveals that the monetization policy amplifies the existing promotional effects of Giveaways campaigns, resulting in both a further increased rating volume and lower rating scores. Table 7 presents the model estimates based on the aforementioned specification. In line with existing literature on promotional effects, we observe that Giveaways participation increases the rating volume of promoted books while lowering their average ratings (see the estimated coefficients β_1 in Columns 1 and 3 of Table 7). Crucially, the coefficient of interest, β_2, indicates that this promotional effect intensifies following monetization: the average rating declines by an additional 0.05 stars post-monetization (as shown in Column 2), and the volume of reviews increases by 14.5 reviews per month (as shown in Column 4). Similar results are found in Section E.3 when focusing on the number and length of text reviews. These findings imply that although book adoption (proxied by rating volume) increases, the participation fee on the supply-side does not lead to improved consumer ratings.

Table 7: Effects of Giveaway Monetization on Average Book Ratings

	Dependent Variable			
	Average Rating		Number of Rating	
Post-Giveaway	-0.201^{***}	-0.184^{***}	11.008^{***}	6.841^{***}
	(0.003)	(0.004)	(0.182)	(0.176)
Post-Giveaway × Post-Monetization		-0.050^{***}		14.454^{***}
		(0.008)		(0.372)
Number of Books	80,498	80,498	81,608	81,608
Overall Mean Rating	3.73	3.73	9.37	9.37
Book Fixed Effects	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes
Observations	961,426	961,426	2,036,383	2,036,383
Adjusted R^2	0.512	0.512	0.131	0.134

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4.3.1 Alternative Specification for Demand Effects: Staggered-Time Difference-in-Differences

The analysis presented in the preceding section implicitly forms a control group using the sample of books involved in the Giveaways program during the pre-monetization period. Consequently, the effect sizes observed post-monetization should be interpreted in relation to an estimated baseline impact of Giveaways campaigns prior to January 2018. In this section, we employ an alternative empirical approach by gathering additional

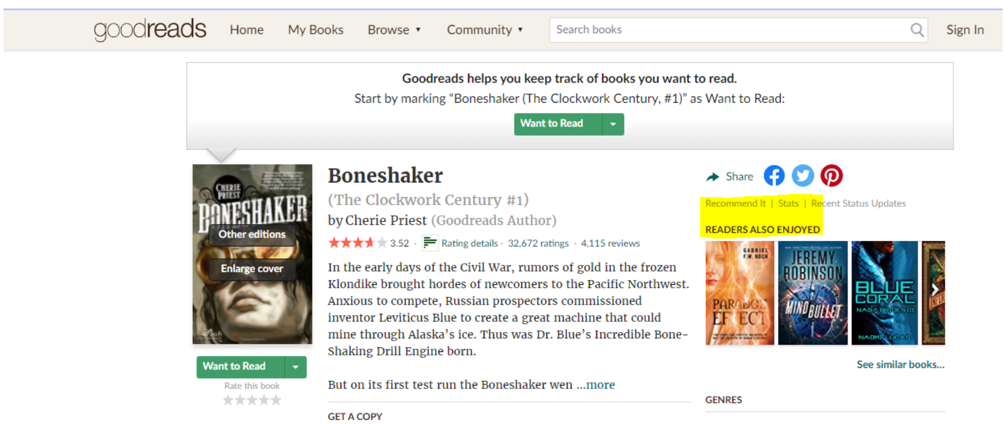
[19]Since we do not directly observe which reviewers receive a free copy of a book from the Giveaways campaign or which reviews are prompted by the program, we interpret the effects on reviews and ratings as intent-to-treat (ITT) effects, i.e., the average impact of paid giveaways on consumer reviews. This approach provides a lower bound on the magnitude of effects, as it incorporates the "noise" from all non-giveaway ratings submitted.

[20]Our unit of analysis is at the year-month level. For the review volume outcome, year-months with no reviews for focal books are set to 0, while for the average rating outcome, these observations are undefined and consequently excluded from regression analysis.

data and explicitly creating a control group comprising books that never participated in the Giveaways program. This method enables us to estimate a different, yet equally relevant, estimand while capitalizing on recent advancements in the econometrics literature concerning staggered-time difference-in-difference, which allows for a more robust construction of counterfactual outcomes for comparison.

Constructing the Control Group Using Goodreads' "Readers Also Enjoyed" Feature To construct this control group, we first identify books that participated in Giveaways, which we will refer to as "focal books". For each focal book, we collect information on the set of books displayed under the "Readers Also Enjoyed" banner (hereinafter referred to as "similar books"), as shown in Figure 10. This section is powered by a data-driven recommender system that showcases books similar to the focal book, based on their characteristics and the frequency with which Goodreads users shelved them together. The similar books and the focal books generally fall within the same category, possess similar characteristics, and are enjoyed by users with similar preferences. To create a suitable comparison group, we adopt a random sampling strategy that associates each similar book with a focal book (each focal book, on the other hand, can be associated with multiple similar books.)[21] This process allows us to create a control group comprising books that did not participate in Giveaways campaigns.

Figure 10: Illustrating of the "Readers Also Enjoyed" Feature



Notes: This is a description of the Goodreads feature that recommends similar books. The similar books and the focal books are typically within the same category and enjoyed by users with similar preferences.

With this control group, we replicate the same pattern of results using the staggered-time difference-in-differences methods proposed by Callaway and Sant'Anna (2021). The primary parameters obtained through this method are group-time average treatment effects, denoted as $ATT(g, t)$. These represent the average treatment effects for units within a specific group (determined by treatment timing) during a particular time period. These parameters serve as a natural extension of the average treatment effect on the treated (ATT), which is identified in the classical case with two periods and two groups and further extended in cases with multiple periods. Group-time average treatment effects also form the foundation for more aggregated treatment effect parameters, such as overall treatment effects or event-study-type estimands. These methods have gained popularity in empirical research where treatment time varies among participating units (see, for

[21] This method enables us to allocate a placebo "treatment date" to each similar book, based on its corresponding focal book, which we will utilize for our triple differences analysis in Section E.8.

example, Bekkerman et al., 2021).

Many applications of these methods involve deploying a single treatment across different units at varying times for distinct cohorts. Our study deviates slightly, as we examine two indicator variables for each book over time: the implementation of the monetization policy in January 2018 and the specific participation time of its Giveaways campaign. To address this challenge, we devise customized estimation strategies to investigate the impact of Giveaways monetization. We conduct the group-time average treatment effect on the treated (ATT) estimation separately for both pre- and post-monetization samples, treating Giveaways participation time as the treatment. Comparing these sets of estimates for the group-time treatment effect under distinct platform policies enables us to assess whether the promotional effect of Giveaways campaigns varies before and after monetization[22]. For the technical details about our customized estimation strategy, please refer to Section E.4.

Employing this estimation strategy, we consistently find that the monetization of the Giveaways program amplifies its promotional effects. The aggregated group-time effects, using similar books as controls, are presented in Table 8. We continue to observe the prevalence of promotional effects for both book samples: average ratings decrease while rating volume increases post-Giveaways participation for books involved in Giveaways campaigns, both before and after the implementation of the new policy. However, the magnitude of these effects varies considerably. The overall treatment effect on rating volume is 5.75 for the pre-monetization sample, whereas it is 15.53 for the post-monetization sample—a threefold increase. Likewise, for star ratings, there is a difference of -0.08 stars between the pre- and post-monetization samples, suggesting that the new monetization policy of the Giveaways program did not enhance word-of-mouth in terms of review valence.

These insights become even more evident when we disaggregate the estimates and examine the dynamic effects over time. We calculate the aggregated treatment effect by length of exposure, which can be interpreted as an event-study style estimator. Figure 11 illustrates our model estimates for these dynamic effects, with 95% confidence intervals adjusted for multiple hypothesis testing. For both outcome variables, the figure reveals systematic differences in effects for books that participate in the Giveaways program before and after its monetization. These differences in effect estimates are statistically significant and align with our prior analysis.

Table 8: Aggregated Group-Time Effects Considering Similar Books as a Control

	ATT	Std. Error	UCB	LCB
Rating Volume				
Pre-Monetization	5.75	0.117	5.517	5.9761
Post-Monetization	15.5344	0.2637	15.0175	16.0514
Average Rating Score				
Pre-Monetization	-0.4417	0.005	-0.4514	-0.432
Post-Monetization	-0.5226	0.0084	-0.539	-0.5061

[22] In an alternative estimation strategy, we examine the full sample and calculate treatment effects by calendar month. This approach allows us to identify heterogeneity based on when a book participates in Giveaways (i.e., before or after January 2018). The results of this model, which are qualitatively similar, can be found in Section E.5

4.4 Mechanism: Increase in Consumer-Book Mismatch

Our analysis has demonstrated that average book ratings decrease following participation in Giveaways, accompanied by an increase in the volume of reviews, with these effects amplified post-monetization. In this section, we explore the potential mechanisms underlying these dynamics. Prior research suggests that products acquired through promotions often receive more negative ratings due to a mismatch between consumer preferences and the product's features (Liu et al., 2019; Zegners, 2019). We investigate whether this pattern of consumer-product mismatch also applies to our context and if Giveaways monetization exacerbates the mismatch. Specifically, we analyze the text of the reviews to quantify fit and assess whether there are more negative reviews stemming from "horizontal" fit mismatch in the post-intervention period. Furthermore, we adopt a dispersion-based metric (Zegners, 2019) as an alternative measure of fit mismatch. Our findings reveal that both measures—negative reviews due to horizontal fit mismatch and rating dispersion—increase following monetization. This evidence supports the notion that Giveaways monetization contribute to a greater consumer-book fit mismatch.

4.4.1 Quantifying Mismatches Using Review Text

We employ the classification method similar to Banerjee et al. (2021), in which we create a hand-labeled training dataset to identify negative reviews arising from a mismatch between the product's fit and the customer's expectations, rather than quality issues. To accomplish this, we enlist two independent workers on Amazon Mechanical Turk to assign each of 3,000 reviews in a sub-sample a discrete score between 1 and 5. In instances of tied scores, we consult a third worker. Subsequently, we aggregate these scores into a binary 0/1 classification for our classifier, excluding neutral scores of 3, which represent only 3% of the data points. Our analysis indicates that approximately 65% of negative reviews were attributable to poor fit, while the remaining 35% result from poor quality. Interestingly, this distribution contradicts the pattern found by Banerjee et al. (2021), who focused on electronics and consumer durables. It is plausible that subjective reviews are inherently more common in domains like books. The labeling prompt we used on MTurk can be found in Appendix E, Figure Figure 20.

We subsequently utilize this labeled data to fine-tune a pre-trained BERT model (Devlin et al., 2018), which has demonstrated superior performance in natural language understanding. A pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a broad range of tasks. Essentially, BERT facilitates bidirectional learning from text by masking (hiding) a word in a sentence and requiring the model to use the words on either side of the concealed word to predict the masked word. With this model, we achieve an F-1 score of 0.78 on the validation sample. We also test other classification models (such as long short-term memory) and simpler tools like support vector machines (results reported in Appendix E, Section E.7), but we ultimately select BERT due to its superior performance in our context.

Ultimately, using this classifier, we assign labels to all the negative reviews in our dataset. We then estimate the same specification as before, with the dependent variable now being an indicator for fit-related negative reviews. If the Giveaways program, and monetization in particular, intensify fit mismatch, we would expect an increase in the proportion of fit-related negative reviews. Indeed, Table 9 shows that the proportion of fit-related negative reviews grows by 15% (0.006 divided by the baseline proportion of fit-related negative reviews, which is 0.04) post-Giveaways, and this increases by a further 12.5% post-monetization. Hence, we find evidence in favor of an exacerbated consumer-book mismatch.

Table 9: Effects of Giveaway Monetization on Proportion of Fit-related Negative Reviews

	Dependent Variable:	
	Proportion of Fit-Related Negative Reviews	
	(1)	(2)
Post-Giveaway	0.008***	0.006***
	(0.0005)	(0.001)
Post-Giveaway × Post-Monetization		0.005***
		(0.001)
Number of Books	80498	80498
Overall Mean Fit Proportion	0.04	0.04
Book Fixed Effects	Yes	Yes
Year-Month Fixed Effects	Yes	Yes
Observations	961,426	961,426
Adjusted R^2	0.125	0.125

Note: $p<0.1$;**p<0.05; ***p<0.01

4.4.2 Quantifying Mismatches Using Rating Dispersion

Next, we show that rating dispersion increases after Giveaways participation, an effect that is further amplified post-monetization. Rating dispersion is commonly used as a measure to indicate preference mismatch (see, e.g., Zegners, 2019). To construct this dispersion measure, for each book, we first compute its mean rating separately before and after participating in Giveaways. We then subtract the corresponding mean from every individual rating received by the book and take the absolute value of the deviation.[23] Subsequently, we estimate

$$|r_{ijt} - \bar{r_i}| = \alpha_j + \gamma_t + \beta_1 \times \text{Post-Giveaway}_{ijt} + \beta_1 \times \text{Time Since Release}_{ijt} + \epsilon_{jt}. \quad (5)$$

We perform this analysis separately for books participating in Giveaways pre- versus post-monetization to determine whether the degree of dispersion differs between these two samples. We find that rating dispersion at the book level increases post-Giveaways, and this effect is further intensified after monetization (0.098 versus 0.081, Table 10). This finding lends additional support to our hypothesis that consumer-book matches deteriorate as a consequence of monetization.

In summary, our findings from both the review text metrics and the dispersion metrics indicate that monetization can influence the efficiency of consumer-book matching, beyond the typical negative reviews caused by mismatches in promotional contexts. The matching process is likely to be hindered when the long tail of genres contracts due to supply concentration, resulting in users with niche tastes struggling to find an appropriate fit.

A potential concern with the above interpretation is that reviewer characteristics and composition may systematically differ post-Giveaways; for example, more critical consumers could be prompted to leave reviews after participating. While this could be a contributing factor, it would not negate the fact that Giveaways negatively impact book ratings, since this selection of reviewers is also driven by Giveaways themselves.

[23] In other words, all ratings posted before monetization are demeaned using the average constructed with pre-Giveaway ratings, and all others are demeaned using the average constructed with post-Giveaway ratings. Similar results are obtained even when we use pre-ratings or overall ratings as the baseline.

Table 10: Effects of Giveaway Monetization on Rating Dispersions

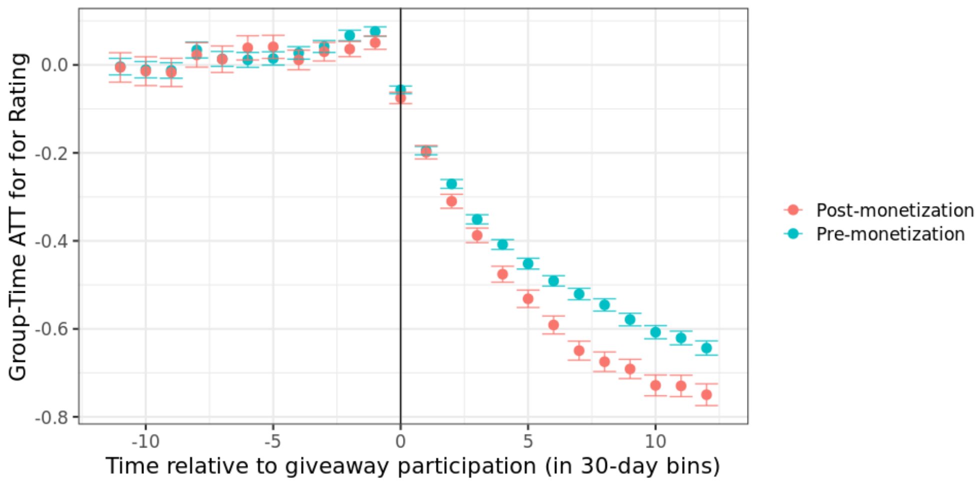
	Dependent Variable:	
	Demeaned Ratings	
	(1: Pre-Monetization)	(2: Post-Monetization)
Post-Giveaway	0.081***	0.098***
	(0.003)	(0.005)
Time since Release	0.0002***	0.0003***
	(0.00003)	(0.0001)
Number of Books	59,760	21,848
Book Fixed Effects	Yes	Yes
Year-Month Fixed Effects	Yes	Yes
Observations	23,800,326	12,830,912
Adjusted R^2	0.081	0.079

Note: $p<0.1$;$^{**}p<0.05$; $^{***}p<0.01$

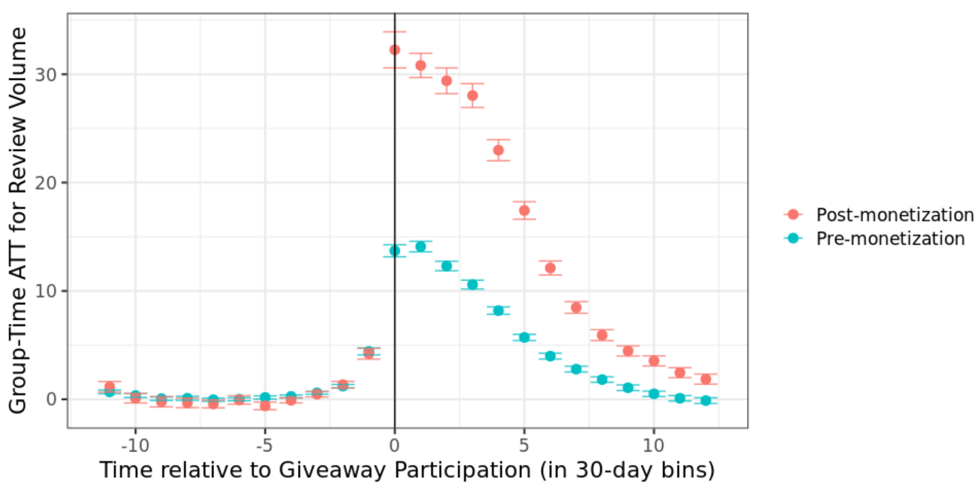
Additionally, it is not clear why this selection effect would be amplified due to monetization. Moreover, if it were solely a selection effect, we would not anticipate the number of reviewers to change. Instead, the observed pattern aligns more closely with our theory that Giveaways, in general, cause more users to adopt the book but lead to poorer evaluations of the product.

A second concern might be that the quality of books participating in Giveaways post-monetization is systematically higher than those participating before. This would imply that their mean ratings post-Giveaways do not change significantly, leading us to estimate a negative difference-in-differences effect when using pre-monetization books as the comparison group. This would also suggest that the monetization policy is pricing out only lower-quality participants, which may actually be in Goodreads' best interest. To address this concern, we refer to Figure 12. As illustrated in Figure 12a, even though the authors participating in Giveaways after monetization are more popular or experienced, there is no evidence of a change in their quality - the overall ratings of participating authors are not significantly different pre- versus post-monetization. The same holds true for the average rating of participating books (Figure 12b). Thus, a stark difference in the quality of participating books is not apparent, at least based on Goodreads ratings, which counters the explanation that monetization is only weeding out low-quality players. Finally, we also demonstrate that our main results are robust to alternate comparison groups in a triple differences specification (see Appendix E, Section E.8).

Figure 11: Dynamic Effects Analysis with "Similar Books" as a Control



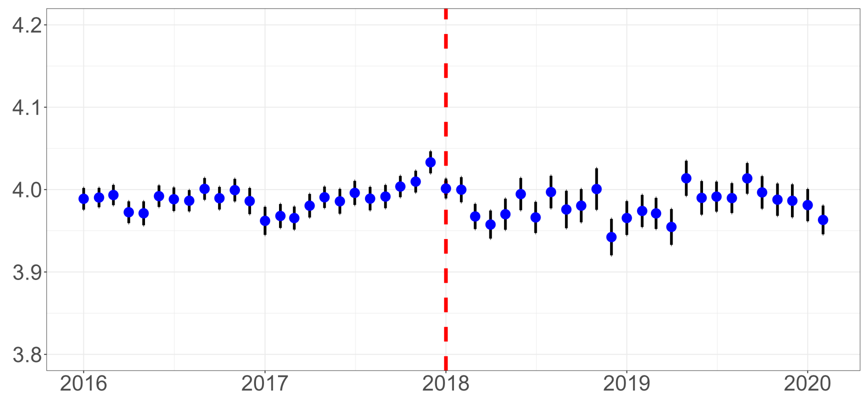
(a) Effect of Giveaways on the Average Rating of Books Participating, Pre- and Post-Monetization



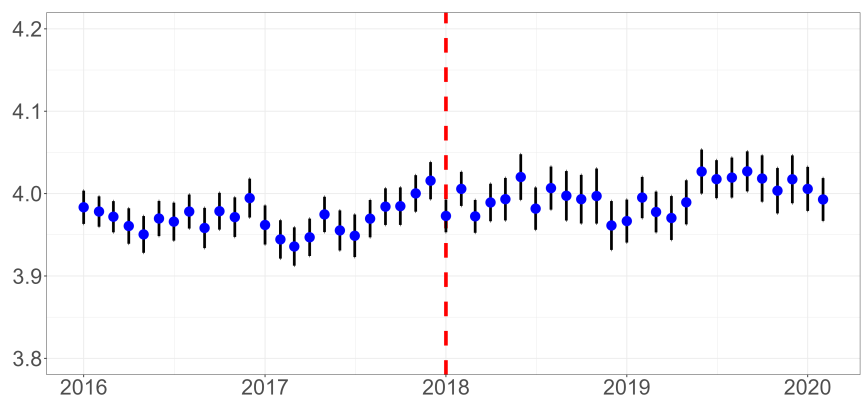
(b) Effect of Giveaways on the Review Volume of Books Participating, Pre-Monetization and Post-Monetization

Notes: For both review volume and valence outcome variables, the figure reveals systematic differences in effects for books that participate in the Giveaways program before and after its monetization. The black vertical line identifies January 2018, and the bars indicate the 95% confidence interval.

Figure 12: No Evidence of Quality Difference in the Giveaways Program



(a) Average Rating of Participating Authors at the Year-Month Level



(b) Average Rating of Participating Books at the Year-Month Level

Notes: We observe no significant change in authors' average ratings and books' average ratings before and after Giveaways monetization. The red vertical dashed line identifies January 2018, and black vertical line segments indicates the 95% confidence interval.

5 Conclusion and Discussions

In this paper, we employ a natural experiment to examine the effects of Goodreads' monetization of its book discovery advertising tool (Goodreads Giveaways) on various stakeholders in the marketplace. First, we observe a shift in publisher representation towards established entities post-monetization, with the Big Five publishers increasing their presence while indie and self-publishers decrease. Moreover, we discover a disproportionate decline in participation from less established/experienced authors as well as female authors. The increased supply concentration also impacts the diversity of book genres available through the program, as more popular genres benefit at the expense of less popular ones. On the demand side, we present evidence that the promotional effects of the Giveaways program intensify after monetization: although books experience a higher adoption rate and consequently receive more reviews, these reviews are, on average, more negative. Lastly, we explore the primary mechanism driving these outcomes and uncover evidence of increased consumer-book mismatches in the post-monetization period, based on an analysis of both review text and rating dispersion. In summary, we reveal that the monetization of Goodreads' advertising tool carries implications for the power distribution among publishers and authors, as well as for the diversity of book genres and the nature of consumer-product match.

Our study presents a more nuanced and comprehensive understanding of monetization consequences, going beyond traditional frameworks. The existing literature and public discourse on platform monetization primarily emphasize network effects, which relate to the platform's value being dependent on the number of users. In our context, concerns about network effects arise when the supply side of the market (authors and publishers) contracts due to the introduction of participation fees, subsequently resulting in diminished value for the demand side (readers) and a potential collapse of the two-sided market. By analyzing the distribution and composition of various participants and products in the marketplace (authors, publishers, books, and readers), we reveal previously unanticipated consequences of monetizing the Giveaways program. Our findings offer valuable insights for understanding the complex factors that should be considered when assessing the success of monetization strategies and the overall health of the ecosystem. These insights indicate that platforms should develop more adaptive incentives to counteract adverse effects and address the diverse needs of different stakeholders within the ecosystem.

These findings stimulate discussion on various nuanced network effects that platforms should take into account. For example, the reduced number of authors and publishers participating in the Giveaways program after monetization could potentially result in a less competitive marketplace for book providers, illustrating a positive same-side network effect. However, the findings also suggest a negative cross-side network effect due to decreased reader participation. Likewise, this monetization policy presents both positive and negative network effects for readers. On one hand, it makes it more difficult for readers to discover a suitable read, as fewer unique books are offered through the program (negative network effect); on the other hand, it raises the likelihood of each reader winning a free book, due to the combination of a larger number of available free books and fewer requests for them (positive effect). Given the complexity and opposing directions of network effects from different perspectives, the effects we observe are not immediately apparent. Furthermore, the question remains as to what specific monetization guidelines can produce an optimal outcome for the development of creative products.

Our research also contributes to the discussion of how platforms can influence fairness within industries. Can a platform reduce unfairness among different players within an industry? To some extent, the answer is yes. Consider a small slipper seller from a modest city, a small independent movie production studio, or a driver who can only offer rides for two hours per day – it used to be challenging and expensive for them to reach their consumers, audiences, and passengers due to their limited resources. The emergence of major platforms (e.g., Amazon, Netflix, Uber, etc.) has transformed the situation, enabling these smaller players to compete with larger companies. However, this research highlights how the strategies a platform employs can disproportionately affect marginal players and place them at a competitive disadvantage against larger entities.

Drawing on our empirical findings, we offer several managerial guidelines for platforms in this context. First, when monetizing their services, platforms may consider subsidizing or providing preferential treatment to small players or under-resourced users in their ecosystem. Although it is common practice for a platform to subsidize large players or star users in its marketplace[24], we are beginning to see smaller players receiving subsidies from platforms. In June 2022, Google offered to subsidize small app developers on its Play store for $90 million[25]. Our findings corroborate this approach — the decreased involvement of marginal or niche players can undermine the diversity of products accessible on a platform, subsequently affecting consumer favorability.

Second, from a platform perspective, an effective monetization policy must strike a balance between direct monetary benefits and other non-monetary aspects crucial to the platform ecosystem's functioning and long-term growth. Based on our back-of-the-envelope calculation, the supply-side participation fees generate approximately $5 to $8 million in revenue within the first two years after implementation. Nonetheless, the platform must also take into account other non-monetary consequences of the Giveaways program monetization. For instance, one type of indirect cost is the loss of word-of-mouth promotion for books created through the program. The introduction of the participation fee results in diminished overall engagement: fewer Giveaways campaigns (and consequently fewer promoted books) from authors and publishers, as well as a decrease in book requests from readers. Each month, roughly 2,000 Giveaways campaigns and 1.5 million book requests vanish following the policy change. For the platform, the program's value, apart from the participation fee, lies in its capacity to generate early buzz about a book, thereby sparking readers' interest in reading and reviewing it. It can be argued that Giveaways campaigns create value for the platform even without monetary contributions from authors and publishers. Similar reasoning can be applied to other non-monetary aspects of the ecosystem, such as the desire for higher representation of female authors in promoted books or, more generally, the promotion of a diverse and inclusive marketplace.

A limitation of our study is that we focus only on one monetization strategy, specifically fixed participation costs paid by providers. It is conceivable that these effects may not arise in the context of other pricing strategies employed by platforms. In fact, as noted earlier, we believe our observed effects could be alleviated by adopting more flexible incentive structures in platform monetization, such as multi-tiered pricing or subsidy schemes for under-resourced platform participants. We consider our contribution to be the introduction of the nuanced and multifaceted nature of platform monetization into the discussion. Another limitation is that we do not investigate market-level equilibrium outcomes or activities on related platforms. It is possible that authors are "substituting" Goodreads for other online communities, like Twitter or book-related subreddits, in their book promotion efforts. However, if true, this actually reinforces our argument that the Giveaways program monetization is counterproductive for the platform, as it drives some community members away from the platform ecosystem. Additionally, given Goodreads' longstanding reputation as the premier destination for book discovery and its extensive user base, it provides a unique value proposition for

[24] For example, as of February 2023, YouTube plans to incentivize more short-form videos with an ad revenue sharing scheme for top creators. See this link: https://www.zdnet.com/article/youtube-short-creators-win-big-with-these-new-updates/, and Spotify paid star podcaster Joe Rogan $200 million to host his show exclusively on its platform.

[25] https://www.ign.com/articles/google-small-app-developers-fund-90-million-settlement

authors to reach a wide audience. No other platform offers comparable opportunities for authors to promote their books as effectively, which means authors and publishers, especially indie authors and those of niche genre books, have limited options for multi-homing or substitution.

Although we focus on the context of Goodreads for book promotion, our findings can be extended to various other contexts. As numerous platforms initially launch as free services (without a revenue stream) and later devise ways to monetize the activities they facilitate (e.g., Facebook, YouTube, LinkedIn, Spotify), our study highlights the importance of evaluating platform monetization from a multidimensional perspective and being attentive to the potential differential impacts on different stakeholders. Our results are particularly relevant for platforms dealing with cultural goods, where consumers may have horizontally differentiated preferences. Additionally, we contribute to the expanding discourse on the market power of digital platforms by illustrating how specific monetization schemes can marginalize smaller or independent authors and publishers who contribute to a diverse range of product offerings.

References

Adamopoulos, P., Ghose, A., and Todri, V. (2018). The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research*, 29(3):612–640.

Aguiar, L. and Waldfogel, J. (2018). Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy*, 126(2):492–524.

Aguiar, L. and Waldfogel, J. (2021). Platforms, power, and promotion: Evidence from spotify playlists. *Journal of Industrial Economics*, 69(3):653–691.

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., and Lalmas, M. (2020). Algorithmic effects on the diversity of consumption on spotify. In *Proceedings of The Web Conference 2020*, WWW '20, pages 2155–2165, New York, NY, USA. Association for Computing Machinery.

Banerjee, S., Dellarocas, C., and Zervas, G. (2021). Interacting user-generated content technologies: How questions and answers affect consumer reviews. *Journal of Marketing Research*, 58(4):742–761.

Bekkerman, R., Cohen, M. C., Kung, E., Maiden, J., and Proserpio, D. (2021). The effect of short-term rentals on residential investment. *Available at SSRN*.

Boudreau, K. J. and Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. *Platforms, markets and innovation*.

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

Byers, J. W., Mitzenmacher, M., and Zervas, G. (2012a). Daily deals: prediction, social diffusion, and reputational ramifications. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 543–552, New York, NY, USA. Association for Computing Machinery.

Byers, J. W., Mitzenmacher, M., and Zervas, G. (2012b). The groupon effect on yelp ratings: a root cause analysis. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 248–265, New York, NY, USA. Association for Computing Machinery.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chen, H., Hu, Y. J., and Smith, M. D. (2019). The impact of e-book distribution on print sales: Analysis of a natural experiment. *Management Science*, 65(1):19–31.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Cowell, F. A. (2000). Chapter 2 measurement of inequality. In *Handbook of Income Distribution*, volume 1, pages 87–166. Elsevier.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Farronato, C. (2017). Pricing mechanisms in online markets. Technical report, Working paper, Harvard University.

Fleder, D. and Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science*, 55(5):697–712.

Foerderer, J. (2020). Interfirm exchange and innovation in platform ecosystems: Evidence from apple's worldwide developers conference. *Management Science*, 66(10):4772–4787.

Foerderer, J., Kude, T., Mithas, S., and Heinzl, A. (2018). Does platform owner's entry crowd out innovation? evidence from google photos. *Information Systems Research*, 29(2):444–460.

Foerderer, J., Lueker, N., and Heinzl, A. (2021). And the winner is ...? the desirable and undesirable effects of platform awards. *Information Systems Research*.

Gini, C. (1921). Measurement of inequality of incomes. *The economic journal*, 31(121):124–126.

Grazian, D. (2005). A digital revolution? a reassessment of new media and cultural production in the digital age. *The Annals of the American Academy of Political and Social Science*, 597(1):209–222.

Guinaudeau, B., Munger, K., and Votta, F. (2022). Fifteen seconds of fame: Tiktok and the supply side of social video. *Computational Communication Research*, 4(2):463–485.

Hagiu, A. (2006). Pricing and commitment by two-sided platforms. *Rand Journal of Economics*, 37(3):720–737.

Hansen, C., Mehrotra, R., Hansen, C., Brost, B., Maystre, L., and Lalmas, M. (2021). Shifting consumption towards diverse content on music streaming platforms. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, pages 238–246, New York, NY, USA. Association for Computing Machinery.

Hirshman, A. (1964). The paternity of an index. 54(5):761–762.

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., and Aral, S. (2020). The Engagement-Diversity connection: Evidence from a field experiment on spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, pages 75–76, New York, NY, USA. Association for Computing Machinery.

Huang, N., Burtch, G., Gu, B., Hong, Y., Liang, C., Wang, K., Fu, D., and Yang, B. (2019). Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science*, 65(1):327–345.

Liu, Y., Lambrecht, A., and Deng, Y. (2019). Price promotions and online product evaluations. *Available at SSRN 3506009*.

Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Nagaraj, A. and Ranganathan, A. (2022). Singing your own praises: Digital cultural production and gender inequality.

Nagaraj, A. and Reimers, I. (2021). Digitization and the demand for physical works: Evidence from the google books project.

Nielsen, R. K. and Fletcher, R. (2022). Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the uk. *Journal of Quantitative Description: Digital Media*, 2.

Parker, G., Dartmouth College, Van Alstyne, M., Jiang, X., Boston University, and Quinnipiac University (2017). Platform ecosystems: How developers invert the firm. *MIS Quarterly*, 41(1):255–266.

Parker, G. and Van Alstyne, M. (2018). Innovation, openness, and platform control. *Management Science*, 64(7):3015–3032.

Parker, G. G. and Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*.

Parker, G. G., Van Alstyne, M. W., and Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company.

Peukert, C. (2019). The next wave of digital technological change and the cultural industries. *Journal of Cultural Economics*, 43(2):189–210.

Peukert, C. and Reimers, I. (2022). Digitization, prediction, and market efficiency: Evidence from book publishing deals. *Management Science*.

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5):645–665.

Rietveld, J., Schilling, M. A., and Bellavitis, C. (2019). Platform strategy: Managing ecosystem value through selective promotion of complements. *Organization Science*, 30(6):1232–1251.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Sharma, S., Telang, R., and Zentner, A. (2021). The impact of digitization on print book sales: Analysis using genre exposure heterogeneity.

Tiwana, A., Konsynski, B., and Bush, A. A. (2010). Research Commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4):675–687.

Toop, D. (1995). Ocean of sound: aether talk, ambient sound and imaginary worlds.

Waldfogel, J. (2017). How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives*.

Waldfogel, J. and Reimers, I. (2015). Storming the gatekeepers: Digital disintermediation in the market for books. *Information Economics and Policy*, 31:47–58.

Waller, I. and Anderson, A. (2019). Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference*, WWW '19, pages 1954–1964, New York, NY, USA. Association for Computing Machinery.

Wu, Y. and Zhu, F. (2022). Competition, contracts, and creativity: Evidence from novel writing in a platform market. *Management Science*.

Zegners, D. (2019). Building an online reputation with free content.

Zhu, F. and Liu, Q. (2018). Competing with complementors: An empirical look at amazon.com. *Strategic Management Journal*, 39(10):2618–2642.

Appendix

A Word-of-mouth on Twitter

For our word-of-mouth analysis in Section 3.3, we use Twitter's new API v2 endpoint of full-archive tweet counts allows us to retrieve the volume of tweet for a given query (official documentation of the endpoint: https://developer.twitter.com/en/docs/twitter-api/tweets/counts). In addition, we further collect the full set of tweets that mentioned both "Goodreads" and "Giveaways" and summarize relevant tweet characteristics of this sample. Doing so, we find that not only does the total amount of word-of-mouth on Twitter drop (as shown in Figure 4), other aspects of Twitter engagement also decrease. We find that fewer unique users are talking about the Giveaway program and also lesser amplification (i.e. retweets) for those tweets, as shown in Figure 13. This unintended consequence is not desirable from the standpoint of the platform, which promotes its main marketing tool, i.e. Giveaways, around having the power to create word-of-mouth.

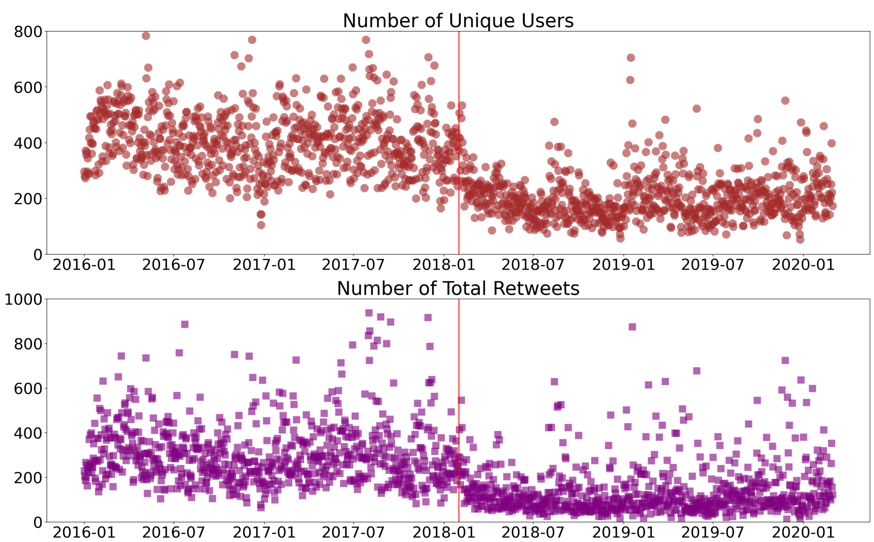


Figure 13: Twitter Engagement Decreases After Monetization.

B Goodreads Ratings Data and its Representativeness

A limitation of collecting ratings and reviews data on Goodreads.com is that the full set of historical ratings and reviews are not visible at the same time - at the time of data collection, users could access at most 10 pages of reviews, with 30 reviews displayed on each page. To circumvent this, we utilize the 'Filter' option and obtain a different set of reviews for each filter, thus enabling us to collect more than 30 * 10 = 300 reviews. The detailed algorithm of data collection is as follows:

1. Navigate to a book URL from the provided list.
2. Scroll to "community reviews"
3. Capture the following aggregate fields: average rating, total ratings, total reviews (i.e., ratings with text).

4. Click on "more filters" and capture the rating distribution

5. Check if the total number of ratings is <=300. If so, we do not need to apply additional filters. Just scraping the 10 pages of reviews displayed will give us the full set of reviews.

6. However, if the total number of ratings are >300 but <=1500, we should use star rating filters:

   - "Go to more filters"
   - Click on a star rating
   - Scrape 10 pages of reviews as in Step 5.

7. Go back to Step 6b to specify the next star rating. Repeat this for 1,2,3,4 and 5 stars one at a time. Thus, the maximum number of pages to scrape is 5 star ratings levels * 10 pages = 50; and the maximum number of obtained reviews is 50*30 = 1500

8. If total number of >1500, we should use star rating AND order filters:

   - "Go to more filters"
   - Click on a star rating
   - Click on "sort order"→"oldest first"
   - Scrape 10 pages of reviews
   - Click on "sort order"→"newest first"
   - Scrape 10 pages of reviews as in Step 5
   - Go back to Step 6b to specify the next star rating. Repeat this for 1,2,3,4 and 5 stars one at a time.

Even after using the above procedure, only a maximum of 3000 ratings and reviews may be obtained. However, there are only about 3% of books that have more than 3000 ratings in our sample. Also, we compare our collected sample to the raw numbers of 1-5 star ratings obtained via book-level metadata, and find very similar distributions, thus providing evidence that our sample is a valid characterisation of books' review distributions.

Goodreads does enable us to see the summary statistics of submitted ratings, i.e, the absolute number of 1,2,3,4 and 5 stars that a book has accumulated (Figure 14). By comparing these to the distribution of ratings captured by our scraper, we can determine the representativeness of our sample. We eventually find qualitatively similar numeric ratings in our context, as shown in Figure 15.

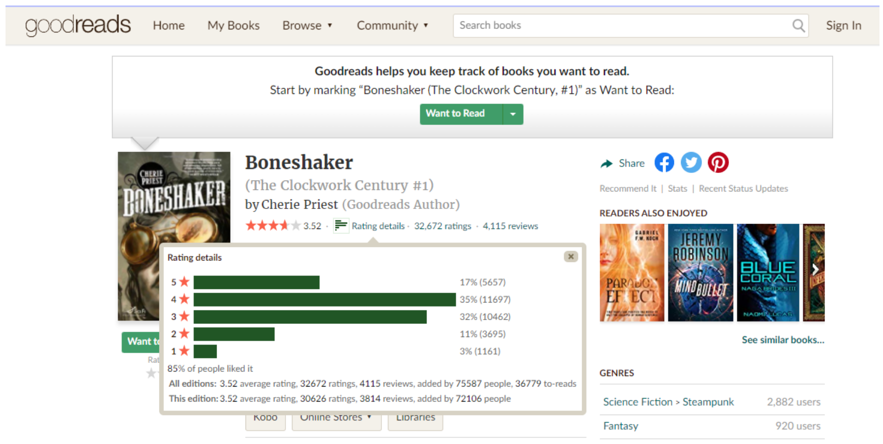


Figure 14: Summary Statistics Available from Goodreads on Review Distributions of a Given book.

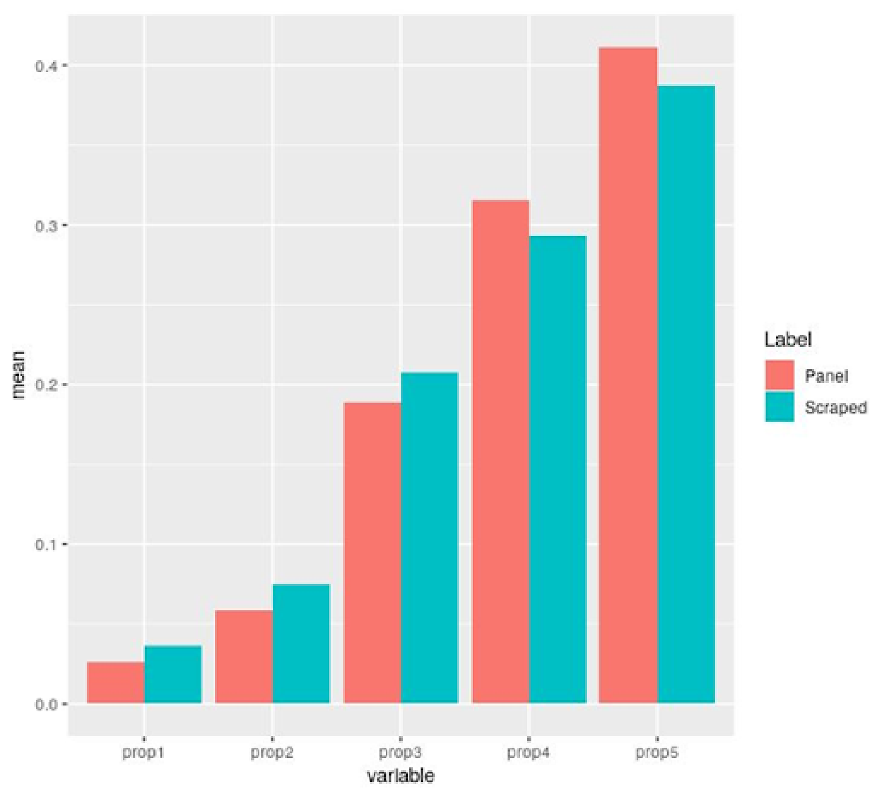


Figure 15: Comparison of Rating Proportions Obtained for Each Star Level from the Scraped Data vs the Goodreads 'Panel' of Summary Statistics: We see that the true distribution very closely tracks our collected sample.

C Publisher Level Analysis

C.1 Publisher Classification

We have 16,326 publishers in the raw data. Since the publisher information is self-reported when a book is listed for a giveaway, publisher information may be subtly different across books even if they are published by the same publisher. For example, two books published by Random House may have listed their publisher as "Random House" or "random house" (cases differ), or Alan Simon Books may be listed as "Alan Simon Books" or "Alan Simon". Therefore, we perform the following pre-processing steps to clean the publisher variable so that books published by the same publisher can be clubbed together:

- Lower the case for all publisher names.
- Remove words that may be omitted in some cases while exist for other cases. For example, books published by lulu.com may have the publisher information as "lulu" or "lulu.com". We remove words/phrases like "publishing", "group", "books", ".com" etc.
- Remove punctuations, white space (some inputs have more than one white space between words).

The above cleaning process leaves us with 15,322 unique publishers.

Moreover, Big Five publishing houses have several subsidiaries under their umbrellas. To identify if a given publisher in our dataset belongs to the Big Five, we do the following:

- Collect all the divisions/imprints of the Big Five Publishing Houses (321 in total).
- Pre-process the division/imprint information as before: lower the case, remove potentially omitted words, remove punctuation/white space.
- Create a dummy equal to 1 if a publisher is one of the 321 imprints/divisions.

Self-publishing entities are identified and tagged in a similar way based on a manual assessment of the book market.[26] There are thirteen self-publishers in our data, including: authorhouse, bookbaby, blurb, createspace, draft2digital, fillinselfpublisher, ingramspark, kindledirect, lulu, streetlib, smashwords, and reedsy after pre-processing.

C.2 Big Five vs Self Publishers: Absolute Numbers

As further evidence of the disproportionate impact of Giveaways monetization, we plot the absolute and relative numbers of Giveaway participation of books published by both the Big Five and self-publishing houses (Figure 16); while we see a drop for both groups, a comparison of the two bar charts indicates a much larger impact of monetization on self-publishing books, both in absolute and relative terms. Importantly, the proportion of books by the Big Five increased sharply from less than 20% to more than 30%, and the proportion of self-publishing books decreased immediately after monetization. This exploration suggests that monetization leads to supply concentration.

[26] Drawing from this list, among others: https://en.wikipedia.org/wiki/List_of_self-publishing_companies, https://blog.reedsy.com/best-self-publishing-companies/, and https://selfpublishing.com/self-publishing-companies/.

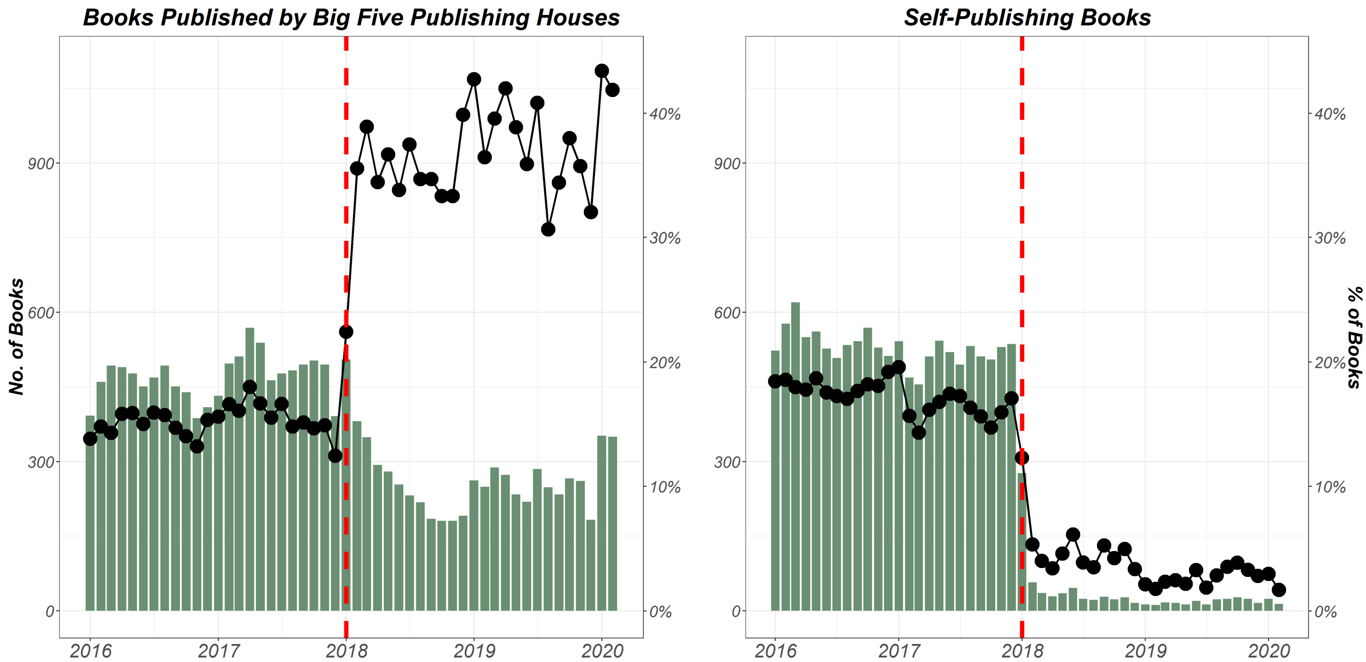


Figure 16: The Absolute Numbers and Proportion of Giveaways from Big 5 vs Self-Publishers.

C.3 Book Level Analysis of Publisher Proportions

In this section, we examine how the mix of books on the market changes in terms of publisher characteristics before vs after monetization. Again, we are interested in the Big Five publishers and self-publishing service providers as our primary examples. We estimate two regressions of the form:

$$P_{jt} = \beta_0 + \beta_1 \times \text{Post Monetization}_t + \delta_t + \epsilon_{jt} \tag{6}$$

where P_{jt} indicates whether a given book j in giveaway month t comes from a Big Five publisher or is self-published respectively. We also add fixed effects for genre and month. Results are reported in Table 11. We see that for a given book, the likelihood that it will be from a Big Five publisher rises by 11% following monetization. Conversely, the likelihood that it will be from a self-publishing house decreases by 5%, thus backing our claims.

Table 11: The book level impact of monetization on publisher proportions.

	Dependent variable:			
	Big Five Book		Self-Publishing Book	
	(1)	(2)	(3)	(4)
Post-Monetization	0.13^{***}	0.11^{***}	-0.07^{***}	-0.05^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Month Fixed Effect	Yes	Yes	Yes	Yes
Genre Fixed Effect	No	Yes	No	Yes
Observations	95,864	95,864	95,864	95,864
R^2	0.02	0.06	0.01	0.06

Note: $p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

D Book Genre Analysis

For the analysis in Section 4.2, we infer book genres based on the most popular shelves they were put on. We took up to the top two bookshelves for each book as its genres after some manual processing of the raw bookshelves data, e.g.,: 1. remove irrelevant shelves, e.g. currently-reading, kindle, and so on; 2. merge some bookshelves that have the same meaning but with different names, e.g. 'historical' and 'history', 'sci-fi' and 'science-fiction'. 3. Take the top 50 bookshelves across all books as the genres categories we will use.

D.1 Entropy Regressions

Here, we use the following fixed effects model with standardized entropy as the dependent variable:

$$\text{Entropy}_t = \beta_1 \times \text{Post Monetization}_t + \beta_2 \times \text{Time Trend}_t + \gamma_m + \epsilon_t \tag{7}$$

The regression includes an indicator for post monetization, which equals to 1 if the observation is after monetization and 0 otherwise. The regression also includes a linear term for time trend to account for common temporal trend across genres as well as month fixed effects to account for seasonality. The results are shown in Table 12. The estimated coefficient for the post monetization indicator (β_1) is negative and significant, indicating a large reduction in genre diversity of books participating in Giveaways after monetization vs before. Note that the time trend does account for part of the reduction in genre diversity; nonetheless, the effect of monetization is still statistically significant and large in magnitude after accounting for that. Based on model estimates in column (2), genre diversity decreases by as much as 1.13 standard deviation after monetization.

Table 12: Impact of Monetization on Genre Diversity

	Dependent variable:	
	Standardized Entropy	
	(1)	(2)
Post Monetization	-1.866^{***}	-1.134^{***}
	(0.134)	(0.280)
Time Trend		-0.028^{***}
		(0.010)
Month Fixed Effect	Yes	Yes
Observations	47	47
R^2	0.853	0.883

Note: $p<0.1$;**p<0.05; ***p<0.01

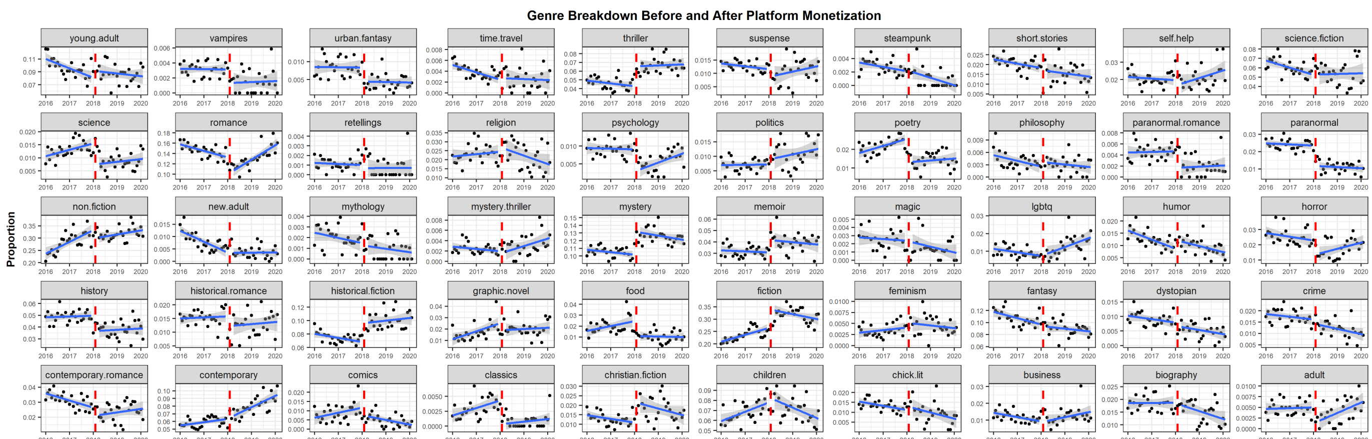


Figure 17: Heterogeneous impact on book genres for all 50 genres

E Effect of Giveaways on Demand

E.1 Analysis Using Full Sample

Our main results restrict the estimation sample to a 2 year window of ratings before and after Giveaways participation. We find below that our main results on ratings, review volume, text reviews as well as review length continue to hold if we look at the full sample (Table 13 and Table 14).

Table 13: The Impact of Giveaways and Monetization on Average Ratings, Measured At the Year-Month Level.

	Dependent variable:	
	Aggregated Avg Rating	
	(1)	(2)
Post Giveaway	-0.455^{***}	-0.363^{***}
	(0.005)	(0.006)
Post Giveaway × Post Monetization		-0.194^{***}
		(0.010)
Number of books	81608	81608
Overall mean rating	3.58	3.58
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2,544,641	2,544,641
Adjusted R^2	0.444	0.445

Note: $p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 14: The Effect of Giveaways and Monetization on Review Volume, Measured at the Year-Month Level.

	Dependent variable:	
	Num. of Rating	
	(1)	(2)
Post Giveaway	4.127^{***}	2.343^{***}
	(0.082)	(0.083)
Post Giveaway × Post Monetization		11.039^{***}
		(0.134)
Number of books	81608	81608
Mean rating count	2.5	2.5
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	12,649,240	12,649,240
Adjusted R^2	0.066	0.078

Note: $p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

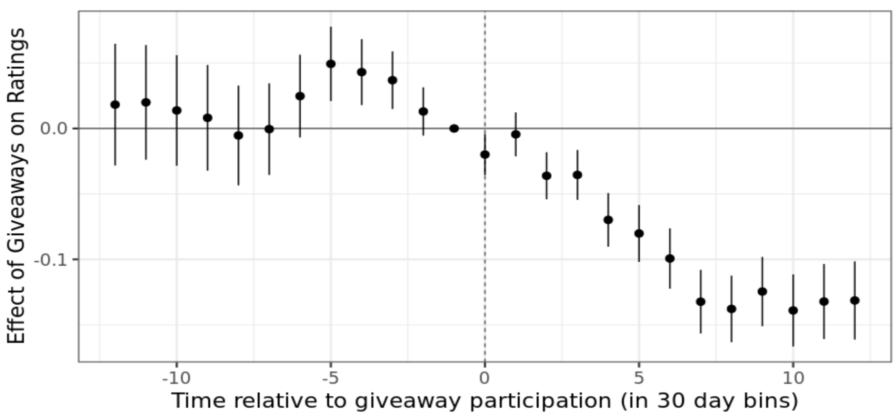
E.2 Main effects over time: average ratings and review volume

To understand the dynamic effects of monetization on ratings and review volume, we estimate:

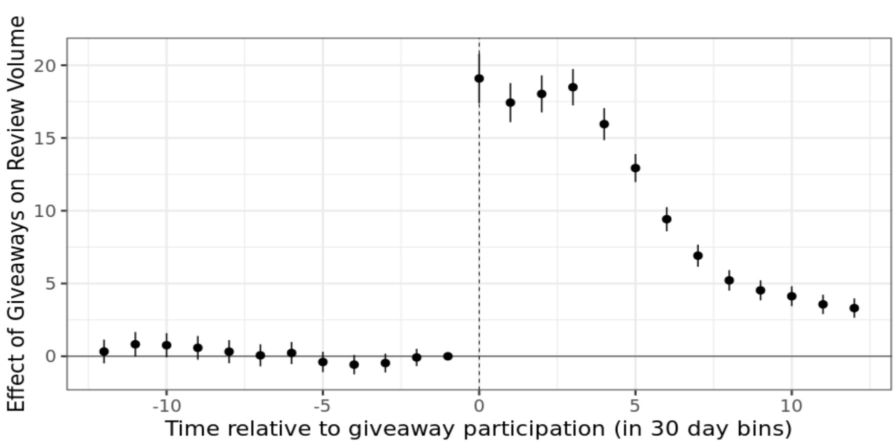
$$y_{jt} = \text{Post-Giveaway} + \text{Post-Monetization} + \delta_t + \sum_{k=-12}^{12} \beta_k \times \mathbf{I}\{D_{jt} = k\}_j \times \text{Post-Monetization} + \epsilon_{jt} \quad (8)$$

where y_{jt} is respectively the average rating or the number of reviews and $\mathbf{I}\{D_{jt} = k\}$ is an indicator for month $k \in -12, 12$ for each book j. Month -1 is taken as the reference level (similar to papers using similar strategies such as Proserpio and Zervas (2017)).

In Figure 18a and Figure 18b, we plot the β_k coefficients from the rating and review volume regressions, and observe no violation of parallel trends in the pre-period.



(a) Effect Over Time of Monetization on Average Ratings: The points plot the β_k coefficient estimates from Equation 8, and the bars indicate the 95% confidence interval.



(b) Effect Over Time of Monetization on Review Volume: The points plot the β_k coefficient estimates from Equation 8, and the bars indicate the 95% confidence interval.

Figure 18: Parallel trends analyses.

E.3 Impact of Monetization on Text Reviews

In addition to book ratings, we also examine the textual reviews of promoted books to quantify whether the number of text reviews and their length change before and after the monetization of Giveaways program. Similar to the results for star ratings, Table 15 shows that the number of text reviews further increases by 3.6 reviews per book post-monetization above and beyond the baseline promotional effect of Giveaways campaigns, while their length did not change significantly. However, on the full sample, review length post-monetization does tend to be shorter by about 14 characters after monetization (Table 16). Taken together, these results suggest that the monetization of Giveaways program facilitate higher product adoption and online word-of-mouth, but did not lead to better consumer satisfaction.

Table 15: Effect of Giveaway and Monetization on Text Review Volume, Measured at the Year-Month Level for 12 Months before and after Giveaways Participation.

	Dependent Variable			
	Number of Reviews		Review Length	
Post-Giveaway	2.847***	1.713***	−116.905***	−120.372***
	(0.094)	(0.107)	(3.801)	(4.854)
Post-Giveaway × Post-Monetization		3.550***		9.759
		(0.199)		(7.727)
Number of Books	81,058	81,058	70,090	70,090
Overall Mean Text Reviews	4.14	4.14	738.54	738.54
Book Fixed Effects	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,427,841	1,427,841	550,024	550,024
Adjusted R^2	0.177	0.176	0.151	0.151

Note: $p<0.1$;**p<0.05; ***p<0.01

Table 16: The Impact of Giveaways and Monetization on Review Length, Measured at the Year-Month Level for the Full Sample.

	Dependent variable:	
	Review length	
	(1)	(2)
Post Giveaway	−228.156***	−222.561***
	(3.286)	(4.115)
Post Giveaway × Post Monetization		−14.149**
		(6.614)
Number of books	70090	70090
Overall mean rating	738.54	738.54
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	1,168,186	1,168,186
Adjusted R^2	0.115	0.115

Note: $p<0.1$;**p<0.05; ***p<0.01

E.4 Technical details of Estimation Strategies using Goodreads' "Readers Also Enjoyed" Feature

To analyse the effect of monetization on customer evaluations, we also make use of an alternate control group and adapt estimation strategies introduced by Callaway and Sant'Anna (2021). Let i denote a book and t a time period (month). Y_{it} is the observed outcome of interest, which is either the volume of rating or average rating score in our setting. For each i, G_{ig} is a binary variable equal to 1 if i first participated in a Giveaways campaign at time g. Finally, $Y_{it}(0)$ denotes i's potential outcome at time t if untreated at time t, and $Y_{it}(g)$ denote i's potential outcome at time t if i was first treated at time g. This framework thus allows for heterogeneous treatment effects with respect to the treatment date. We set the anticipation period to equal 1 to account for the fact that authors and publishers may adjust their campaign strategies and hence affect rating metrics up to one month before Giveaways participation, so that $Y_{it}(0) = Y_{it}(g)$ for all $t < g - 1$. We also assume that once a unit is treated, it remains treated and that the intensity of the treatment is the same for all units. The relation between observed and potential outcomes is as follows:

$$Y_{it} = Y_{it}(0) + \sum_{g=1}^{T} [Y_{it}(g) - Y_{it}(0)] G_{ig}. \tag{9}$$

We then define the group-time treatment effect as

$$ATT(g,t) = E[Y_{it}(g) - Y_{it}(0) | G_{ig} = 1]. \tag{10}$$

Given that different potential outcomes cannot be observed for the same unit at the same time, we define our causal effect in terms of the similar book sample, which constitutes a "never-treated" group.[27]

Once these group-time averaged treatment effects are computed, we can choose to aggregate them into a simple average estimate as follows:

$$TE_o = \frac{1}{\kappa_o} \sum_{g} \sum_{t>g} \omega_g ATT(g,t), \tag{12}$$

where ω_g is proportional to the number of books with treatment date g and κ_o normalizes the weights so that they sum up to one (results in Table 8). We implement the estimator in R using the package *did*, developed by Brantly Callaway and Pedro Sant'Anna. We estimate TE_o for average ratings score and rating volume on two separate samples: one for books that participate in the Giveaways program before its monetization and one for books participate after monetization. This implicitly take into account the problem sample selection and comparability and identifies our main effects under a different set of assumptions. Namely, within the pre- and post-monetization periods, we compare our focal books (that participate in giveaways) to similar books in the same period.

For dynamic effects, we define TE_e as the weighted average of ATT(g,t) for all t and g such that $t-g=e$;

[27] The main identifying assumption for this analysis to be valid is the conditional parallel trends assumption. In short, for each $g \in G$ and $t \in 2, ..., T$ such that $t \geq g - \delta$ (δ being the anticipation period, which equals 1 in our case),

$$E\left[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1\right] = E\left[Y_t(0) - Y_{t-1}(0) \mid X, C = 1\right], \tag{11}$$

where C denotes units that do not participate in treatment during any time period (the similar books in our case; for more details on the assumptions, see Callaway and Sant'Anna, 2021). This assumption requires that if treated groups had instead not been treated, then their outcome would evolve in the same way as groups that have never been treated. This cannot be directly tested because the left-hand side of Equation 11 is not observed. However, we can test whether groups of observations have a different trend with respect to the amount of time left until Giveaways participation in the pre-treatment period; we test this in Figure 11.

that is,

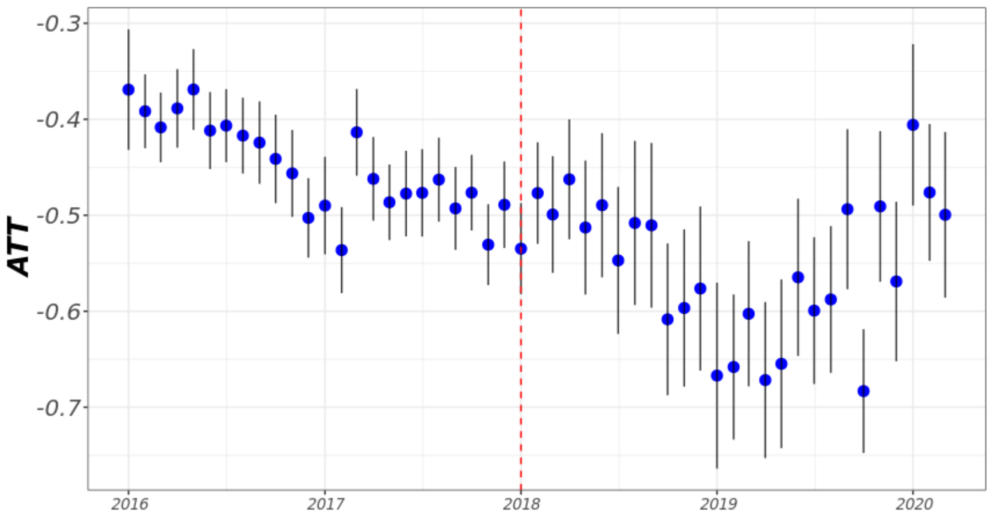
$$TE_e = \frac{1}{\kappa_e} \sum_g \omega_g ATT(g, g+e), \tag{13}$$

where, again, ω_g is proportional to the number of books with treatment date g and κ_e ensures the weights sum up to one. The parameter TE_e can be interpreted as an event study. More precisely, for $e < 0$, TE_e captures the trend in outcomes for groups that are e periods away from giveaway participation relative to other groups that never participate. Conversely, for $e > 0$, TE_e captures the trend in outcomes for groups that are e periods post-Giveaways relative to groups that never participate. Figure 11 reports our estimates (and 95% confidence intervals adjusted for multiple hypothesis testing). The figure shows flat pre-trends and a significant difference post-giveaway between both samples, both for the review volume and the average ratings.

E.5 Heterogeneous effects of Giveaways participation over time

The method of Callaway and Sant'Anna (2021) also allows us to explicitly compute treatment effects by cohort. For instance, their methodology allows the treatment effect for a book that participates in Giveaways in 2017 to vary from that of one participating in 2019. Since we want to show that there are indeed heterogeneous treatment effects in our setting, we also conduct a separate analysis on the full sample of focal books along with their similar books. This time, we focus on aggregating ATT(g, t) at the calendar month level, which tells us the effect for books that participate in Giveaways for a particular calendar month. In Figure 19, we indeed find that the treatment effect of participating in Giveaways is higher for review volume and lower for ratings for the months that follow monetization (January 2018).[28] This lends further convergent evidence that monetization exacerbates the promotional Groupon effect. These effects are further confirmed by a triple differences design using the same control group data, available in Section E.8.

[28]The final few months of 2020 appear to yield a U-shaped relationship for the average rating, but note that these estimates are also noisier due to fewer post-monetization ratings being observed.

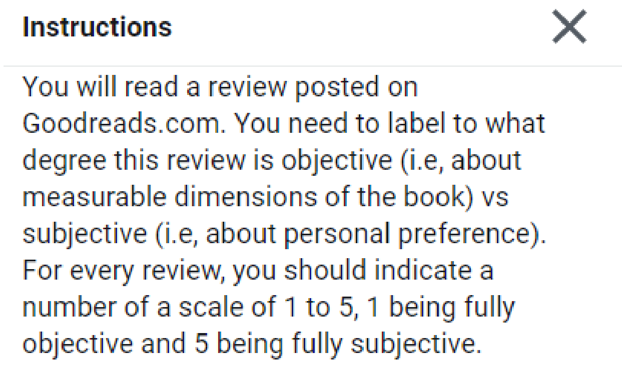


(a) Effect of Giveaways on the Average Rating of Books Participating in Giveaways in a Given Calendar Month

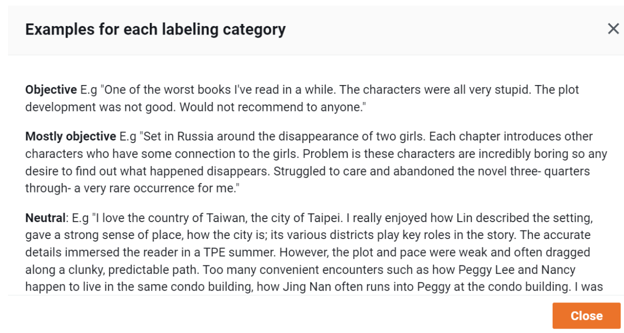
(b) Effect of Giveaways on Review Volume of Books Participating in Giveaways in a Given Calendar Month

Figure 19: Heterogeneous Treatment Effects of Giveaways Participation over Time.

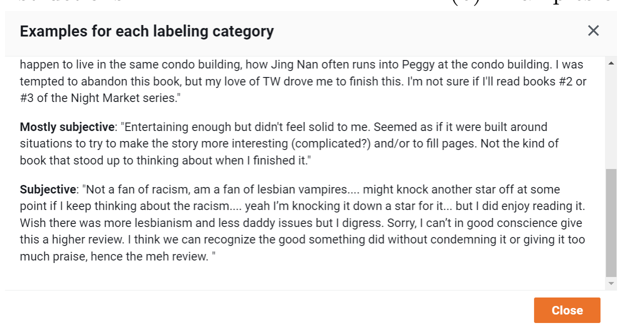
E.6 Detailed pre-processing steps for BERT



(a) General instructions.



(b) Examples of objective reviews.



(c) Examples of subjective reviews.

Figure 20: Training Data Classification Task on MTurk.

To understand whether consumer-book mismatch is heightened after the monetization of Giveaways, we look at the text of reviews and build a classifier using manually labeled data collected on MTurk (instructions in Figure 20). We undertake the following pre-processing steps to run our BERT classification model:

1. Since fit related reviews dominate the training data, we first up-sample the quality reviews to account for imbalance.

2. Next, we use a BERT tokenizer to convert the raw text into a format that BERT recognizes. In particular, this involves (i) padding the total length of reviews (ii) adding specific tokens for BERT (iii) adding attention masks to ensure that the attention layers do not consider the padded tokens when learning context.[29]

3. Next, we split the data into training and validation sets (keeping 25% for validation and the rest for training).

4. We choose an AdamW optimizer because it performs better than Adam by preventing overfitting and having an improved implementation of weight decay.

5. Finally, we create a schedule with a learning rate that decreases linearly (Loshchilov and Hutter, 2017)

[29]The tokenizer we use comes from the 'bert-base-uncased' model in the Python HuggingFace library.

E.7 Alternative classifier for fit reviews

As a robustness check, we also train a linear C-Support Vector Machine (C-SVM) classifier (Cortes and Vapnik, 1995) on the set of manually labeled reviews. To perform the classification task, we remove common stopwords, and then tokenize and stem the text of each negative review into a bag-of-words representation, thus obtaining word frequencies for each negative review. We use these word frequencies as predictors to train a classifier that predicts whether a negative review arises primarily from poor fit or poor quality.[30]

We train our classifier on an 80% random sample of our labelled data, holding out the remaining 20% to evaluate the classifier's performance. The C-SVM classifier has one tunable parameter, C, which intuitively calibrates the trade off between classification accuracy and having a larger-margin separating hyperplane. We select a value for C using 5-fold grid search cross-validation. We evaluate the performance of the classifier with an F-1 score (which combines the precision and recall of a classifier into a single metric by taking their harmonic mean). We find an F-1 score of 63% based on 5-fold cross validation.[31]

Finally, using this classifier, we get labels for all the negative reviews in our data. We estimate the same specification as before, but with the dependent variable being an indicator for fit-related negative reviews. If giveaways in general, and monetization in particular exacerbates fit mismatch, we would expect there to be an increase in the proportion of fit related negative reviews. Indeed, in Table 17, we see that the proportion of fit related negative reviews go up by 4.5% post-giveaways, and this increases by a further 1.3% post-monetization. Hence, we find evidence supportive of consumer-book mismatch being heightened. Similar results are also obtained on the full sample, both for SVM (Table 18) and BERT (Table 19)

Table 17: The Impact of Giveaways and Monetization on Proportion of Fit Related Negative Reviews, Measured at the Year-Month Level for 12 Months Before and After Giveaway Participation.

	Dependent variable:	
	Proportion of fit related negative reviews	
	(1)	(2)
Post Giveaway	0.008***	0.007***
	(0.0004)	(0.001)
Post Giveaway × Post Monetization		0.005***
		(0.001)
Number of books	80498	80498
Overall mean fit proportion	0.03	0.03
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	961,426	961,426
Adjusted R^2	0.129	0.129
Note:		*p<0.1; **p<0.05; ***p<0.01

[30] We also considered using bigrams as predictors, but we did not see significant improvement in out-of-sample predictive power.

[31] We also try other supervised and semi-supervised classifiers that yield similar or lower F-1 scores. While a high classification accuracy is warranted, a noisy classifier actually works against us in this setting by adding noise to the dependent variable and biasing any subsequent regression estimates towards zero (attenuation bias). Any effect we find in this scenario will thus be lower bound of the possible effect size, and thus a more conservative test of our hypotheses.

Table 18: The Impact of Giveaways and Monetization on Proportion of Fit Related Negative Reviews, Measured at the Year-Month Level Before and After Giveaway Participation.

	Dependent variable:	
	Proportion of fit related negative reviews	
	(1)	(2)
Post Giveaway	0.009***	0.006***
	(0.0003)	(0.0004)
Post Giveaway × Post Monetization		0.006***
		(0.001)
Number of books	81608	81608
Overall mean fit proportion	0.03	0.03
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2,544,641	2,544,641
Adjusted R^2	0.087	0.087
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 19: The Impact of Giveaways and Monetization on Proportion of Fit Related Negative Reviews, Measured at the Year-Month Level Before and After Giveaway Participation.

	Dependent variable:	
	Proportion of fit related negative reviews	
	(1)	(2)
Post Giveaway	0.006***	0.004***
	(0.0003)	(0.0004)
Post Giveaway × Post Monetization		0.005***
		(0.001)
Number of books	81608	81608
Overall mean fit proportion	0.03	0.03
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	2,544,641	2,544,641
Adjusted R^2	0.082	0.082
Note:		*p<0.1; **p<0.05; ***p<0.01

E.8 Triple differences

As an additional robustness check using a different set of identifying assumptions, we conduct a triple-differences analysis (difference-in-differences-in-differences, or DDD) to estimate the effect of monetization on ratings and review volume, using data collected from the "Readers also Enjoyed" feature of the website. Similar to the analysis described in Section 4.3.1, we can associate every 'focal' book that participates in Giveaways to a set of 'similar books' that do not. Hence, the time of treatment for the similar/control books are defined in terms of the Giveaways participation date of the focal book.

Conceptually, DDD takes place in two DD steps. First, we compute a DD for books that participate in Giveaway after monetization, similar to Equation 4. Then, we adjust this DD for unobserved differences by subtracting from it the DD after monetization for books that do not participate in Giveaways. We then estimate a triple difference specification using matched-pair book fixed effects.

In our particular case, we consider three separate indicator variables: Treated (=1 if the book participates in a giveaway), Post-Giveaway (=1 for all periods after Giveaway participation) and Post-Monetization (=1 for books that participate in Giveaway after January 2018). The estimate of interest for us is the effect of giveaway participation on books that participate in a giveaway after monetization. This is given by β_5 in the equation below:

$$\begin{aligned} r_{jt} = \alpha_j + \gamma_t + \beta_1 \times \text{Post-Giveaway}_{jt} + \beta_2 \times \text{Post-Giveaway}_{jt} \times \text{Treated}_j + \\ \beta_3 \times \text{Post-Monetization}_{jt} \times \text{Treated}_j + \beta_4 \times \text{Post-Giveaway} \times \text{Post-Monetization} + \\ \beta_5 \times \text{Post-Giveaway} \times \text{Post-Monetization} \times \text{Treated}_j + \epsilon_{jt} \end{aligned} \tag{14}$$

The triple difference estimator can be computed as the difference between two difference in-differences estimators. Despite this, the triple difference estimator does not require two parallel trend assumptions to have a causal interpretation. The intuition is that the difference between two biased difference-in-differences estimators will be unbiased as long as the bias is the same in both estimators. In that case, the bias will be differenced out when the triple difference is computed. This requires only one parallel trend assumption, in ratios, to hold.

Estimating this specification, we again find that average ratings of books that participate in Giveaway post-monetization are lower by about 0.75 stars, whereas review volume increases by 23 (Table 20).[32] Hence, even after considering a more stringent functional form that accounts for omitted confounders, our main results hold, and effect sizes are in fact larger.

[32] For these results, we focus on the entire data and not just a 12 month window before vs after monetization, due to the presence of a control group that can account for trend effects. This is what leads to differences in the sample sizes across the tables.

Table 20: The Effect of Giveaways on Average Ratings and Review Volume, Measured at the Year-Mon Level, Estimated Using a Triple-Difference Specification.

	Dependent variable	
	Avg Rating	Review Volume
Post Giveaway	-0.198^{***}	6.03^{***}
	(0.005)	(0.074)
Post Giveaway × Treated	0.212^{***}	-7.74^{***}
	(0.005)	(0.078)
Post Giveaway × Post Monetization	0.194^{***}	-3.61^{***}
	(0.008)	(0.132)
Post Monetization × Treated	0.555^{***}	-0.97^{***}
	(0.016)	(0.036)
Post Giveaway × Post Monetization × Treated	-0.731^{***}	14.27^{***}
	(0.014)	(0.182)
Number of books	81622	81622
Book FEs	Yes	Yes
Year-month FEs	Yes	Yes
Observations	6,516,591	25,302,820
Adjusted R^2	0.324	0.083
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$	