Sherlocking: The Effects of Platform-Owner Entry on the Competitive Behavior of Third-Party Firms*

Benjamin T. Leyden†

July 13, 2026

Platform owners increasingly compete against the firms they host. I study how third-party developers respond to Apple's 2016–2021 entries into 22 App Store submarkets, using same-market, never-treated Google Play apps as controls. Apple's entry deters third-party entry by about 22 percent, with no detectable change in exit, and shifts incumbents toward paid pricing even though Apple's own product is free. These pooled effects mask striking heterogeneity: many markets show no effect, the monetization response is larger for OS-integrated entries, and effects vary with proximity to Apple's product. This heterogeneity suggests tailored oversight rather than categorical restrictions on platform-owner entry.

JEL: L13, L86, L40

1 Introduction

Platform-owner entry into digital marketplaces has become a central issue in antitrust policy and the regulation of digital markets (De Chant, 2022). The practice—where marketplace operators compete with third-party sellers on their own platforms—has prompted significant regulatory debate worldwide. The European Union's Digital Markets Act, enacted in 2022, explicitly restricts "gatekeepers" from engaging in "any form of differentiated or preferential treatment... in favour of products or services it offers itself" (EUR-Lex, 2022). In the United States, the proposed American Innovation and Choice Online Act (first introduced in 2021 and reintroduced in 2026) would similarly bar platforms from using "nonpublic data... to offer... products or services... that compete... with products or services offered by business users" (Congress.gov, 2022, 2026). In some cases, politicians have called for an outright ban on the practice (Warren, 2019). These blanket restrictions reflect growing concerns that platform owners exploit their unique position to discriminate

*I am grateful to Qihong Liu, Chiara Farronato, Devesh Raval, Mike Walker, Laura Lasio, Daniel Ershov, Xuan Teng, and participants at the 20th Annual International Industrial Organization Conference, the NYU Economics of Strategy Workshop, the NUS Economics of Platforms Workshop, ASSA 2023, the US DOJ Economic Analysis Group, the Cornell Strategy and Economics of Digital Markets Conference, the ZEW ICT Conference, HKUST, the 2025 CESifo Economics of Digitization Conference, the University of Sydney, APIOC 2025, the Digital Competition Conference 2026, the 2026 MaCCI Annual Conference, the University of Oklahoma, the TSE Economics of Platforms Seminar, and the WEAI 101st Annual Conference for helpful comments. I thank Yujie Feng for excellent research assistance, and I acknowledge the Cornell Center for Social Sciences for financial support.

†Cornell University; CESifo; leyden@cornell.edu

against competitors, leveraging privileged information and control over marketplace infrastructure to advantage their own offerings (Crémer, de Montjoye, and Schweitzer, 2019; Scott Morton et al., 2019).

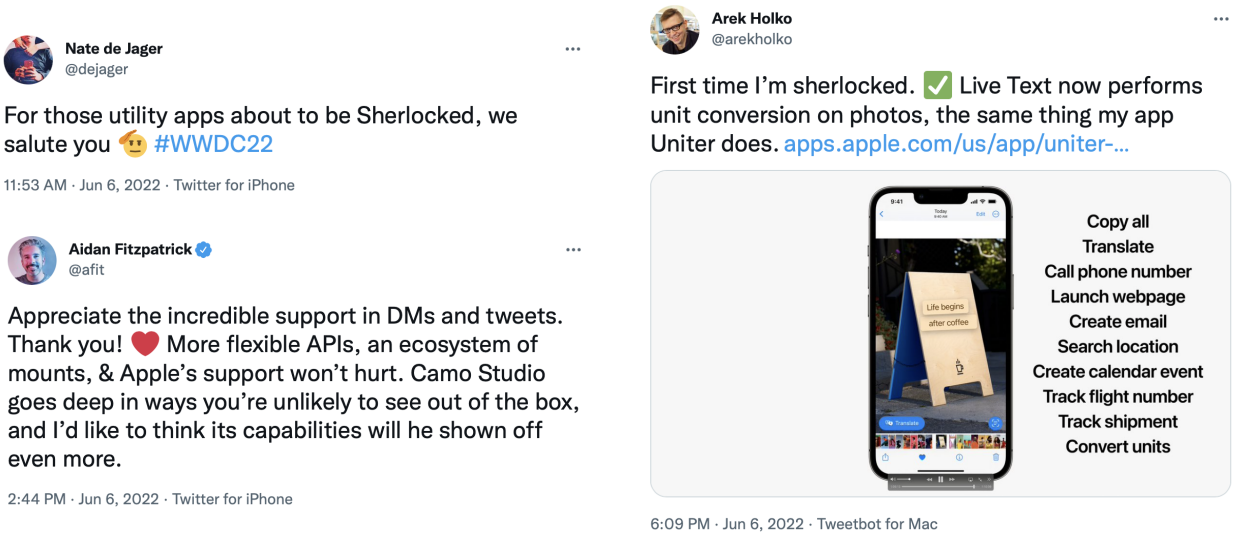
Yet the economic effects of platform-owner entry remain theoretically ambiguous (Anderson and Bedre-Defolie, 2024; Hagiu, Teh, and Wright, 2022). While entry introduces a typically well-resourced competitor that may engage in self-preferencing or restrict access to essential features, it can also generate positive spillovers through technology diffusion, market expansion, and complementary infrastructure investments. The debate over appropriate regulatory responses, ranging from outright bans to conduct remedies to laissez-faire approaches, hinges critically on understanding how large these effects are, how much they vary across markets and competitive contexts, and what mechanisms drive them.

To help inform this debate, I study how third-party developers respond to Apple's release of competing apps and software functionality. This "Sherlocking" phenomenon—named after Apple's incorporation of the third-party search app *Watson*'s functionality into its own *Sherlock* product—is a first-order concern for developers, who routinely track Apple's entry into App Store submarkets (see Figure 1). Developers face substantial uncertainty about whether and when Apple might enter their product space, complicating long-term investment decisions. For those whose products are targeted, the stakes are high: Apple's entry has the potential to render years of development effort obsolete overnight.

In the context of mobile software platforms, platform-owner entry takes two distinct forms that have received insufficient attention in both academic literature and policy discussions. Apple's entry into submarkets within its App Store marketplace occurs through either *standalone entry*—developing and releasing a new separately installable application—or *integrated entry*—incorporating competing functionality directly into the operating system. Standalone entry involves Apple creating a new app that users can discover and choose to download from the App Store, competing alongside third-party offerings. Examples include Measure (an augmented-reality measurement tool), Translate (language translation), and Schoolwork (classroom management). These apps appear in search results, have product pages with descriptions and ratings, and can be installed and uninstalled by users, though some may come pre-installed by default. The competitive dynamic is relatively straightforward: consumers actively choose whether to adopt Apple's offering over alternatives.

Integrated entry involves Apple integrating functionality directly into iOS that automatically becomes available to all users. Examples include health tracking features integrated into the Health app, Apple Music enhancements, Notes app improvements, and new camera features. This functionality cannot be separately uninstalled and often comes bundled with new APIs that third-party developers can leverage, representing a more fundamental alteration of the platform's capabilities. Users receive these features automatically, without making an active adoption decision. The competition is therefore unavoidable: third-party apps must compete with functionality that every user already possesses. Integrated entry can also generate stronger technology spillovers, as OS-level

Figure 1: Developer Reactions to "Sherlocking"



Note: Tweets from developers reacting to Apple's WWDC 2022 announcements of new iOS features that compete with existing third-party apps.

integration often introduces new hardware capabilities or software frameworks that benefit the broader developer ecosystem.

This paper provides comprehensive evidence on platform-owner entry effects by studying 22 instances of Apple entering submarkets in its App Store between September 2016 and April 2021. The analysis encompasses 10 standalone app releases and 12 integrated OS features, enabling systematic comparison across entry types and market characteristics.

My empirical strategy leverages a staggered difference-in-differences design, using comparable Google Play Store apps as controls for App Store apps affected by Apple's entry, with within-market cross-platform comparisons providing the primary source of identification. Market definitions rely on semantic similarity of app descriptions, with apps exceeding a similarity threshold considered part of the competitive market around each entrant.

I make three key contributions to the literature on platform competition and regulation. First, by studying 22 entry events rather than a single case, I formally establish and characterize cross-market heterogeneity in competitive responses. I apply meta-analytic diagnostics to market-specific treatment effects to document that markets frequently disagree on the direction of response and that null effects are widespread across most outcomes. This heterogeneity is itself a substantive finding: it is evidence against an assumption common to the leading regulatory proposals on both sides that entry's competitive effect is uniform enough across markets to be addressed by a single, categorical rule. Second, I develop a novel empirical approach using continuous distance measures derived from text embeddings to capture within-market heterogeneity based on competitive proximity. This methodology recognizes that, in a market of differentiated apps where no two products are perfect substitutes, not all apps in an affected market are equally "treated" by platform entry. Third, I show that entry implementation matters: integrated OS features generate competitive

effects on key pricing margins substantially larger than for standalone app releases.

Pooled estimates show that Apple's entry deters new competitors by 22%, shifts incumbent pricing from free to paid, and increases rating count growth, a proxy for user demand and engagement. These averages, however, are taken over markets that respond very differently. Formal heterogeneity diagnostics confirm that, for most outcomes, sampling variation accounts for little of the cross-market dispersion—the differences across markets are real, not statistical artifacts—and a substantial share of markets show no detectable effect, particularly on quality metrics. The pooled average is thus a genuine summary of the central tendency, not a description of any single market's experience.

When effects do occur, they vary across markets and space. Entry type proves crucial. Integrated OS features generate systematically larger effects than standalone app releases. Price increases for integrated entry are about 12 times as large as for standalone entry, while the proportion of free apps declines by 7.6 percentage points for integrated entry but shows essentially no change for standalone entry. This entry-type gap in pricing appears among both incumbent apps and new entrants. Moreover, the effects that vary with distance are the competitive ones: the gains in ratings and demand accrue to apps that remain differentiated from Apple's product, and they fade, if anything reversing, for the closest competitors. The monetization response has a different shape. The shift from free to paid is broadly distributed across competitive distance, so proximity to Apple's product marks where the gains from entry give out, not where pricing moves.

These patterns of heterogeneity—across markets, between entry types, and within markets by distance—suggest that platform-owner entry effects depend critically on contextual factors that current regulatory proposals largely ignore. The evidence strongly supports targeted, context-specific oversight rather than blanket restrictions.

This paper connects several literatures. The question of how the firms a marketplace hosts respond when its owner enters against them has a long history in the economics of retail, where a retailer's own store brand competes with the national brands on its shelves (Mills, 1995; Bergès-Sennou, Bontems, and Réquillart, 2004). That work finds the incumbent response neither uniform nor obvious. Ward et al. (2002), for instance, document that national brands did *not* cut prices against private-label entry, contrary to the prevailing wisdom.[1]

Platform-owner entry is the digital incarnation of this problem, and its two forms differ in how closely they mirror that precedent. Apple's standalone entry, a separately installable app placed in the App Store alongside third-party offerings, is the close counterpart of a store brand, except that it is typically free rather than priced. Its integrated entry, built directly into the operating system and reaching every user automatically, has no clear offline analogue. These departures from the retail regime, in which a priced product competes for finite shelf space, help explain why the response I document runs opposite to the retail benchmark. Rather than meeting a cheaper rival

[1] A parallel theoretical literature on "encroachment" reaches a similar verdict from the opposite end of the supply chain, studying an upstream supplier that sells directly to consumers in competition with its own retailer. Arya, Mittendorf, and Sappington (2007), for example, show that the encroached-upon retailer can even be made better off.

by holding or cutting price, third-party apps move *toward* paid pricing, as a free first-party entrant annexes the ad-supported tier instead of undercutting a price ladder.

The same digital setting also reshapes how competition can be measured. A boxed grocery good sits in a single aisle, whereas a software product can span several at once, as when one app bundles email, calendar, and task management. Discrete category membership is therefore a poor guide to which apps actually compete, and I measure competitive exposure continuously instead, as an app's distance to the first-party product in the embedding space of their descriptions. And because I observe a broad set of affected submarkets rather than a few categories, I can treat cross-market heterogeneity as an object of estimation.

Closer to my setting is a small empirical literature on the effects of platform-owner entry, in which studies typically examine a limited slice of the competitive response. Förderer et al. (2018) study a single entry event, Google's 2015 release of Google Photos, and find that it increased complementary innovation among Android photography apps, operating through a demand-side attention spillover rather than competitive pressure. Jamison and Yu (2026) study first-party entry across many App Store and Google Play niches and find that it raises third-party app downloads on net, a demand expansion that outweighs any business stealing. Zhu and Liu (2018) turn to Amazon's retail marketplace, where many sellers list the same items, and find that Amazon selectively begins reselling successful third-party products itself—competing head-to-head with the sellers that had offered them—raising those products' demand while discouraging affected sellers from growing on the platform.[2] Wen and Zhu (2019) study the *threat* of Google's entry into Android app markets (proxied by Apple's prior entry into the equivalent iOS markets) and, across three entry events, find that threatened developers cut innovation and raise prices on directly affected apps while redirecting innovation toward unaffected and new apps. Kapacinskaite and Mostajabi (2024), studying Apple's 2017 entry into file management with its Files app, document that the response splits by complementor positioning: developers active only on the App Store intensify effort on their affected apps, while multi-homing developers redirect effort toward Google Play.

A striking feature of this literature is that its findings do not point in a common direction. Entry has been found to raise complementary innovation in one setting and to depress it in another. On the theoretical side, Choi, Kim, and Mukherjee (2025) predict the entry deterrence I document in Section 3: a platform's ability to observe and imitate third-party sellers deters their entry, a threat the platform in turn manages through its referral fee and its information-sharing policy.

My analysis studies 22 separate Apple entry events between September 2016 and April 2021, spanning a range of affected App Store submarkets and estimated within a single, unified cross-platform design. In contrast to a platform reselling the identical good, as in Zhu and Liu (2018), the entry I study introduces a differentiated first-party competitor into a market of imperfect substitutes: no third-party app in my data reaches a cosine similarity of 0.90 to Apple's product, and the median sits near 0.64. That residual differentiation makes price a central margin of response

[2] Jiang, Jerath, and Srinivasan (2011) model this behavior in the same Amazon setting: a platform owner learns from third-party sellers' early sales which "mid-tail" products to begin selling itself, and sellers may strategically mask demand to deter entry.

in this setting. Each incumbent holds a distinct position rather than a perfect substitute, so it retains demand of its own that it can re-sort and re-price when a free first-party product enters.

Observing many distinct entries by one platform owner across a common set of monetization and quality outcomes lets me treat that variation as an object of study in its own right. In addition to a single average effect, I estimate the full distribution of market-level responses, distinguish standalone app releases from integrated OS features, and measure how effects decay with an app's competitive distance from the entrant. While I focus on the competitive effects of entry, complementary work by Halckenhaeusser et al. (2026) examines what drives platform-owner entry decisions, finding that Apple's iOS entries target categories characterized by low user satisfaction, limited innovation, and high market concentration.

A related literature isolates the specific mechanisms of platform advantage that may lie behind it. Studies of self-preferencing document that platforms favor their own products in search and recommendations, with mixed welfare implications (Teng, 2026; Lee and Musolff, 2025; Lam, 2023; Raval, 2023).[3] Work on "insider imitation" studies how platforms use proprietary data to select entry markets (Madsen and Vellodi, 2025; Hagiu, Teh, and Wright, 2022). The approach taken here captures the net effect of all channels through which platform entry affects competition.

This paper also relates to existing work on "platforms as regulators", which recognizes that platform owners shape competitive dynamics through rules, design choices, and technical standards (Boudreau and Hagiu, 2009). Recent work shows how platform policies on privacy (Kesler, 2023; Li and Tsai, 2026; Johnson et al., 2026; Cheyre et al., 2026), ratings systems (Gutt et al., 2019; Leyden, 2025), and product categorization (Ershov, 2024) affect within-platform competition. My findings suggest that platform-entry decisions represent another "regulatory" lever.

Methodologically, the paper advances techniques for defining and measuring competition in digital markets. Building on Hoberg and Phillips (2016) and Leyden (2022), I use transformer-based language models to construct product spaces from text descriptions, enabling more precise market definitions than traditional classification systems. The continuous distance measure allows for gradient effects within markets, reflecting the fact that competitive pressure varies with product differentiation.

The remainder of the paper proceeds as follows. Section 2 describes the data sources, market definition methodology, and sample construction. Section 3 presents the empirical analysis, beginning with pooled average treatment effects before formally quantifying cross-market heterogeneity using meta-analytic diagnostics, examining the role of entry type, and analyzing how competitive proximity shapes responses. Section 4 concludes with a discussion of policy implications and directions for future research.

[3] Leung, Qi, and Strumpf (2026), for instance, show that Netflix favors its own originals over equally popular licensed titles in its Top 10 rankings. That setting differs from mine in what the disadvantaged party can do: a licensed title is content the platform curates centrally, not a firm that can answer back, whereas the App Store developers I study respond to entry by repricing, updating, or repositioning.

2 Data

I study platform-owner entry effects using data on the complete population of apps in Apple's App Store and Google's Play Store over a five-year period. I use similar apps on the Play Store as a control group for App Store apps affected by Apple's entry.[4] The Play Store represents the other major mobile marketplace with a comparable developer ecosystem and user base. The use of the Play Store as a control is particularly compelling because Google did not systematically enter the same markets at the same times as Apple, and both platforms experienced similar technological and demand shocks during the sample period. Below, I describe the data sources, Apple's entry events, the market definition procedure, and the construction of the two analysis samples.

2.1 Data Sources

I use a monthly panel dataset of all apps available in the US App Store and Google Play Store provided by the data analytics company AppMonsta (AppMonsta, 2016). The full sample runs from January 2015 to June 2022 and consists of 14.4 million unique apps across both platforms. For each app-month, I observe product characteristics including name, price, average rating, rating count, and whether the app was updated that month. The data also includes developer-written text descriptions that serve as apps' primary marketing content and provide a reliable indication of product functionality and purpose. Additionally, I use daily ranking lists for product sales and revenue to construct sample restrictions focused on actively competing apps.

Key variables for the analysis include monetization outcomes: the download price (Price), whether the app is free to download (Free), whether the app offers any in-app purchases (IAP), and the app's price conditional on it being non-zero (Price-Paid). I measure product quality using apps' average ratings, which are expressed on a 1–5 star scale (Average Rating), and whether an app was updated by its developer in the given month (Update). As a proxy for consumer demand and engagement, I use a measure of the arrival of new ratings (Δ Log Rating Count).[5] Finally, I measure the frequency of entry and exit using log market-level counts of new app entries and exits.[6] An app is counted as entering in the first month it appears in a market's data and as exiting in the last month it appears before dropping out of the panel.

2.2 Apple's Entrants

I study 22 instances of Apple entering submarkets within its App Store ecosystem between September 2016 and April 2021. Apple implements entry through two distinct strategies. In *standalone entry*, Apple releases a new separately installable application that users can discover and download from the App Store; my sample includes 10 such events. These apps compete directly alongside

[4] Play Store in-app purchase (IAP) data is only available starting September 2016. As a result, 5 markets (Breathe 10.0, Clock 10.0, Health 10.0, Homekit 10.0, and Magnifier 10.0) are excluded from all IAP analyses because they entered before sufficient pre-period data was available.

[5] The rating count growth rate is calculated as $\Delta \ln(\text{Rating Count} + 1)$.

[6] Specifically, $\ln(\text{Entry or Exit Count} + 1)$.

Table 1: Apple Entry Events and App Counts by Platform

			Incumbent Sample		Market Dynamics Sample	
Market	Entry Date	Type	App Store	Play Store	App Store	Play Store
Breathe 10.0	2016-09	Integrated	158	59	470	137
Clock 10.0	2016-09	Integrated	176	113	301	190
Health 10.0	2016-09	Integrated	19	9	22	15
Homekit 10.0	2016-09	Integrated	80	63	145	93
Magnifier 10.0	2016-09	Integrated	55	10	48	6
Camera 11.0	2017-09	Integrated	136	59	201	142
Development 11.0	2017-09	Integrated	220	22	218	49
Notes 11.0	2017-09	Integrated	303	96	315	94
Radio 11.1	2017-10	Integrated	34	16	64	20
Animoji	2017-11	Standalone	68	23	176	79
Health 11.3	2018-03	Integrated	117	67	171	58
Schoolwork	2018-06	Standalone	113	39	116	52
Measure	2018-09	Standalone	61	18	80	26
Walkie Talkie 12.0	2018-09	Integrated	143	74	108	149
Ecg	2018-12	Standalone	37	14	27	13
Apple Music 13.0	2019-09	Integrated	75	40	38	31
Reality Composer	2019-09	Standalone	37	12	41	18
Apple Research	2019-11	Standalone	21	8	34	15
Sleep	2020-09	Standalone	114	58	79	46
Translate	2020-09	Standalone	327	157	206	103
Blood Oxygen	2020-10	Standalone	9	4	20	9
Find Items	2021-04	Standalone	29	19	23	14

Notes: App counts at Apple entry date based on $\theta = 0.6$ threshold. The incumbent sample counts apps present throughout the 18-month balanced panel that ranked in the top 200 for at least 10% of days. The market dynamics sample counts apps entering or exiting during the analysis window that ever ranked in the top 200 for at least 1 day. These are distinct populations: incumbent apps are established apps present before entry, while market dynamics apps are new entrants and exiters.

third-party offerings with their own product pages, ratings, and the ability to be uninstalled, though some may come pre-installed on devices by default. Examples include Schoolwork (classroom management), Measure (augmented-reality measurement), and Translate (language translation). In the second case, *integrated entry*, Apple integrates competing functionality directly into its operating system; my sample includes 12 such events. This functionality is automatically available to all users, cannot be uninstalled, and often introduces new APIs that third-party developers can leverage. Examples include Camera improvements, Health app additions, Apple Music updates, and developer-framework releases such as ARKit and Core ML. For example, Apple's 2017 introduction of ARKit provided third-party developers with the ability to place virtual objects in a user's physical space using the device's camera and motion sensors, without building their own computer-vision systems. Table 1 provides the complete list of all 22 entry events with their dates and types. This broad set of entry events, spanning diverse product categories and entry modes, enables systematic analysis of heterogeneous platform-owner entry effects.

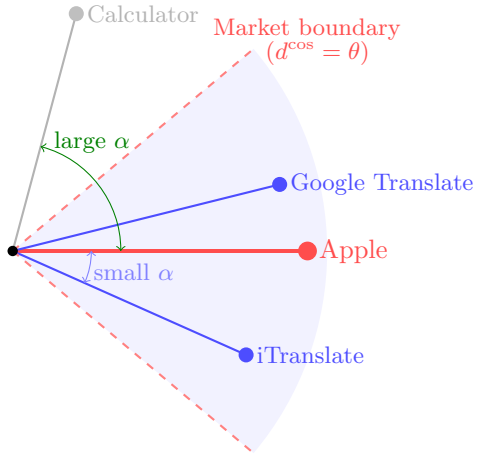
Figure 2: Semantic Similarity in Vector Space: An Illustration

Apple Translate:
"Translate lets you quickly and easily translate your voice and text..."

Google Translate:
"Translate between up to 249 languages...."

iTranslate:
"Seamlessly translate text, websites, objects, or start voice-to-voice conversations in..."

Calculator App:
"Easily type out math with the basic and scientific calculators for quick solutions..."



Note: Left panel: Excerpts from product descriptions used as input to the SBERT model. Right panel: SBERT maps descriptions to vectors in a high-dimensional space where semantic similarity corresponds to geometric proximity (small angles between vectors). Translation apps cluster together despite using different terminology, while the calculator app is distant. The dashed lines represent market boundaries at threshold θ.

2.3 Market Definition

Given the millions of products observed over the sample period, identifying which third-party apps are potentially affected by platform-owner entry requires precise market definitions. Platform-provided categories (e.g., Games, Productivity) are too broad. The Productivity category alone encompasses password managers, shift scheduling apps, document editors, and many other distinct product types.

I define markets using semantic similarity of product descriptions, leveraging recent advances in natural language processing.[7] Specifically, I apply Sentence-BERT (SBERT), a transformer-based neural network model, to embed text descriptions into a 384-dimensional vector space in which semantically similar products are located closer together (Reimers and Gurevych, 2019). These embeddings are trained so that functionally similar apps—regardless of the specific terminology used by the developers—cluster in the embedding space. Markets are defined using the same SBERT process across platforms, ensuring that treated App Store apps and control Play Store apps operate in the same product spaces.

Let J^{AS} and J^{PS} be the sets of apps on the App Store and Play Store, respectively. Each app $j \in J^{\text{AS}} \cup J^{\text{PS}}$ is characterized by a description vector, $desc_j$, constructed using app j's first observed description. For standalone entries, the description vector $desc_a$ used in Equation (1) is the App Store product-page description that was live at the entry date. For integrated entries, which lack natural product pages, I construct comparable descriptions using Apple's iOS release notes as the

[7] There is a growing literature using text documents to characterize markets, following pioneering work by Hoberg and Phillips (2016). See also Hoberg and Phillips (2025), Leyden (2022), Han et al. (2024), and Gugler, Szücs, and Wohak (2024).

source text. These release notes document each feature's functionality at the time of their release. Because release-note entries can be terse, I augment brief descriptions with additional context from Apple's developer documentation. I then generate App Store-style descriptions using Anthropic's Claude Opus 4.7 with few-shot learning, providing a fixed sample of 20 actual descriptions of Apple's standalone apps as style examples to ensure the generated text matches the tone, structure, and vocabulary typical of Apple.[8] This approach produces description vectors that are semantically comparable across standalone entries, integrated entries, and third-party apps, enabling consistent market definition.

I measure the similarity between any app j and Apple entrant a using cosine similarity

$$d^{\cos}(desc_j, desc_a) = \frac{desc_j^\top desc_a}{\|desc_j\|_2 \, \|desc_a\|_2} \in [-1, 1],$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Geometrically, cosine similarity is the cosine of the angle α between the two vectors, so vectors pointing in similar directions (small α) have similarity close to 1. Figure 2 illustrates this approach using language translation apps as an example: SBERT maps textually distinct but functionally similar descriptions to nearby vectors, while unrelated apps like calculators are placed farther away.

Given a first-party entrant a with description vector $desc_a$, I define the market J_a^θ[9] as all apps j satisfying:

$$J_a^\theta = \{j \in J^{\text{AS}} \cup J^{\text{PS}} : d^{\cos}(\text{desc}_j, \text{desc}_a) \geq \theta\}. \tag{1}$$

The parameter θ governs market size: higher values impose stricter similarity requirements, yielding smaller, more tightly defined markets, while lower values produce larger markets that include more distant competitors. In practice, I set $\theta = 0.6$, which defines relatively tight, more focused competitive sets than lower alternative thresholds. This ensures the treated sample consists of apps that are meaningfully similar to the entrant, reducing concerns about including tangentially related products that might bias estimates toward zero. The threshold choice balances two considerations: setting θ too high yields markets that are too narrow and may miss relevant competitors, while setting it too low includes apps that are not meaningful substitutes.[10]

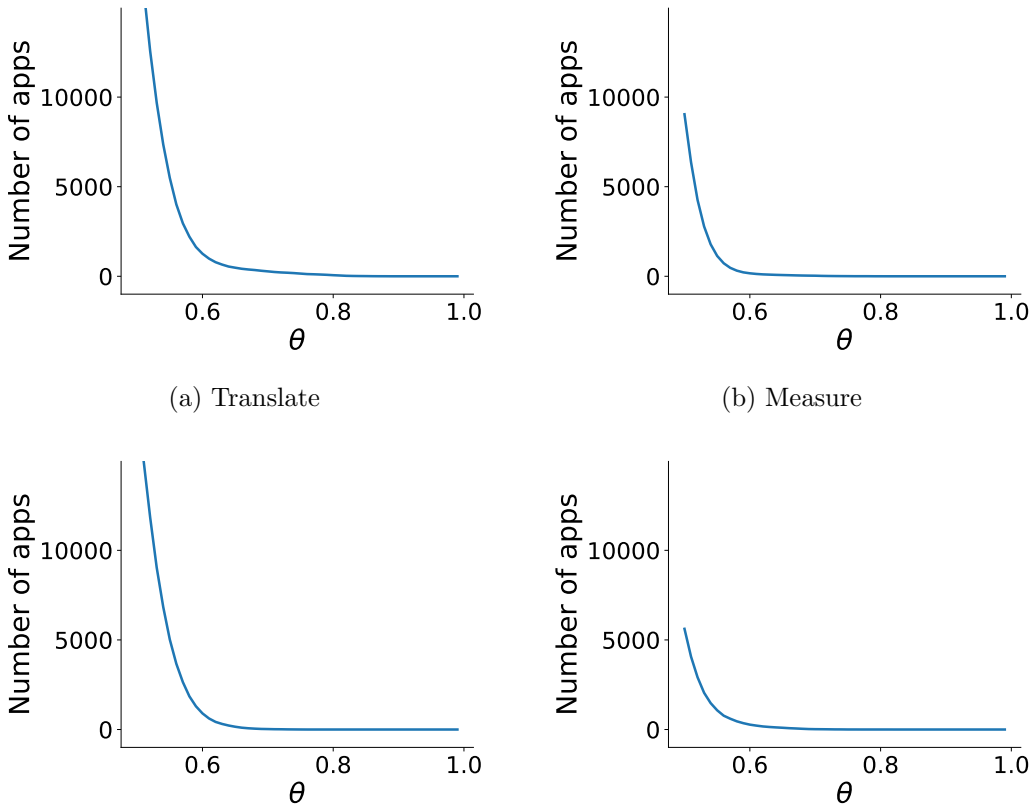
Each app j in market J_a^θ is thus characterized by a continuous similarity to the entrant, $d_j \in [\theta, 1]$, enabling analysis of how treatment effects vary with competitive proximity. In practice, even the closest third-party app falls well short of the entrant's position: no app in my data reaches a cosine similarity of 0.90 to Apple's product, and the median in-sample app sits near 0.64. Thus,

[8] The synthesis pipeline pins the model snapshot and uses a frozen 20-example style block (e.g., Pages, Keynote, GarageBand) chosen for semantic neutrality with respect to the markets being defined. The prompt instructs the model to produce 200–400 word descriptions that focus on user benefits and match the style of the examples. I validate that synthesized and real Apple descriptions land in comparable regions of the SBERT embedding space: across 10 standalone entries, the median cosine similarity between the actual App Store description and a description synthesized from the same product is 0.901 (IQR [0.886, 0.927]), supporting the comparability of standalone and integrated market definitions.

[9] The Apple entrant is not treated as a third-party app and is thus not included in the market samples.

[10] Appendix D.5 shows that estimates for my primary analyses are stable across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$.

Figure 3: Market Size Heterogeneity: Cumulative App Counts by Similarity Threshold



(a) Translate

(b) Measure

(c) Schoolwork

(d) ECG

Note: Each panel plots the cumulative number of apps with cosine similarity to the first-party entrant at or above threshold θ. The first-party entrant is located at $\theta = 1$. As θ increases moving rightward, fewer apps meet the threshold, yielding smaller, more focused markets.

Apple enters as a differentiated alternative within the product space, not a near-copy of an existing app. This continuous treatment approach, presented in Section 3, represents a methodological advance over binary in/out market definitions.

Market sizes vary substantially based on product space density. Figure 3 illustrates this heterogeneity for four representative markets. Each panel shows, among the App Store apps that satisfy the top-200 ranking restriction introduced in the following section, how many fall within cosine similarity θ of the entrant as θ ranges from 0.5 to 1.0. The sharp increases visible at $\theta < 0.6$ demonstrate that loosening the threshold would dramatically expand markets to include many tangentially related apps, motivating the choice of a conservative threshold that focuses on close competitors.

2.4 Analysis Samples

One challenge in analyzing app markets is that many apps are hobby projects, abandoned apps, and experimental products that are not actively competing during the sample period. To address this, and study only competitively relevant products, I restrict my analysis sample based on daily product ranking lists, which rank apps on both platforms based on their number of downloads or the amount of revenue earned.[11]

Specifically, I construct two analysis samples. The *incumbent sample* restricts to apps that ranked 200$^{\text{th}}$ or better on any of these category-level download- or revenue-based ranking lists for at least 10% of days they were available on the platform. This requirement selects for apps with sustained competitive relevance. For each Apple entrant, I construct an 18-month balanced panel centered around the first-party entry date. The resulting dataset contains 59,616 app-months of observations across 22 markets.

For the entry/exit and entering cohort composition analyses in Section 3.1, I use a less restrictive *market dynamics sample* that includes all apps that ever ranked in the top 200 for at least one day. The incumbent sample's stricter restriction selects on sustained post-entry performance, which is appropriate for the incumbent balanced panel but could bias extensive margin analysis: new entrants are mechanically less likely to satisfy a stringent ranking requirement, and conditioning on post-entry ranking risks selecting on outcomes affected by treatment. The market dynamics sample avoids this concern while still excluding apps with no evidence of meaningful market participation. The entering-cohort analysis observes 4,262 apps in the month they enter, across 22 markets. Robustness of incumbent results to alternative sample restrictions, including a pre-treatment-only variant of the incumbent rule, is reported in Appendix D.7.

Table 2 presents summary statistics for the two datasets used in the analysis, comparing pre-treatment (9 months before Apple's entry) and post-treatment (9 months after) averages.[12] Panel A reports app-month level statistics for incumbent apps, defined as those present in the market before Apple's entry, pooling App Store and Play Store apps. The pre-treatment average price of $0.75 reflects that 81.7% of apps are free; among paid apps, prices average $4.09. The post-treatment period shows substantial changes: average prices increase to $1.53, driven by a 17.7 percentage point decline in the share of free apps. In-app purchase adoption increases slightly (from 35.4% to 37.2%), while update frequency declines modestly (from 24.7% to 23.5%). Rating count growth (log-differenced) is positive in both periods, suggesting ongoing user engagement over time.

Panel B reports market-platform-month level log entry and log exit counts. Entry rates decline modestly between periods, while exit rates increase slightly. Panel C reports statistics for the entering cohort analysis, a repeated cross-section where each app is observed once upon entry. Entrants in both periods are predominantly free (88% pre-treatment, 78% post-treatment) with

[11] Revenue ranking lists only reflect revenues earned through the platform—namely, download revenues (for non-free apps) and in-app purchase revenues (which can include subscriptions). Advertising revenue and revenue earned through other sources, such as the developer's website, are not reflected in these lists.

[12] Market-level summary statistics are provided in Appendix A, and Appendix A.4 reports these statistics separately for standalone and integrated entry markets.

Table 2: Summary Statistics

	Pre-Treatment	Post-Treatment	Change
Panel A: Incumbent Apps			
Price	0.748	1.526	+0.777
Price (Paid Apps)	4.090	4.238	+0.148
Free	0.817	0.640	-0.177
In-App Purchases	0.354	0.372	+0.018
Avg. Rating	3.978	3.982	+0.004
Update	0.247	0.235	-0.013
Δ Log # Ratings	0.047	0.030	-0.017
N	29,808	29,808	
Panel B: Market-Level Entry and Exit			
Log Entry	1.499	1.394	-0.105
Log Exit	1.250	1.373	+0.123
N	396	396	
Panel C: Entering Cohorts			
Price	0.372	0.682	+0.309
Price (Paid Apps)	3.152	3.114	-0.038
Free	0.882	0.781	-0.101
In-App Purchases	0.143	0.163	+0.021
Avg. Rating	4.332	4.315	-0.017
Log # Ratings	0.796	0.955	+0.159
Similarity	0.648	0.650	+0.001
N	2,302	1,960	

Notes: Panel A reports app-month level statistics for incumbent apps observed in an 18-month balanced panel around each entry event. Panel B reports market-platform-month level log entry and exit counts. Panel C reports statistics for apps observed once upon entry (repeated cross-section). Pre-Treatment refers to the 9 months before Apple's entry; Post-Treatment refers to the 9 months after.

substantially lower prices than incumbents, consistent with new entrants using zero-price strategies to compete against established apps. Similarity to the Apple entrant is nearly identical across periods (0.65).

3 Empirical Analysis

I estimate how third-party developers respond to platform-owner entry using a staggered difference-in-differences design that exploits variation in the timing of Apple's entry across 22 submarkets, comparing App Store apps to similar apps on the Google Play Store as controls. I examine responses along two margins using the two samples defined in Section 2.4. Extensive-margin responses (entry, exit, and the composition of entering cohorts) are analyzed using the market dynamics sample. Intensive-margin responses (monetization and quality outcomes for established apps) are analyzed using the incumbent sample.

The central aim of this section is to characterize how the effects of entry vary across markets. The pooled averages I first report summarize the typical response, but they are only part of the story: the markets they pool over respond very differently from one another. Some effects are sizable; others are small or statistically indistinguishable from zero—in-app purchases, update frequency, and several engagement margins among them. I find that part of that variation is systematic: the monetization response is larger where Apple enters through OS integration than through a standalone app, and within a market it falls most heavily on the apps closest to Apple's product.

In Section 3.1 I present the pooled average responses, and in Section 3.2 I formally quantify the cross-market heterogeneity behind them. In Section 3.3, I ask whether entry type (integrated versus standalone) explains variation across markets, and in Section 3.4 I document how competitive proximity shapes developers' responses.

The fundamental identification challenge is that one cannot observe counterfactual outcomes: what would have happened to affected apps absent Apple's entry. The identifying assumption is that absent Apple's entry, App Store and Play Store apps in the same product market would have followed similar outcome trajectories. The credibility of this design relies on two key institutional features: Google did not systematically enter the same markets at the same times as Apple, avoiding contamination of the control group, and both platforms experienced similar technological and demand shocks during the sample period. With regard to the first requirement, Google was a long-prior incumbent, a later follower, or absent altogether in all but one of the cases considered here. I tabulate each case and examine the one exception (ARCore, in the Development 11.0 market) in Appendix A.5. The second requirement is supported by the fact that both platforms run frontier mobile operating systems on comparable hardware and compete for largely the same consumers in the mobile phone market.

Beyond Google's own entry decisions, a second threat to the control group is spillover among multi-homing apps, products available on both platforms. If a developer adjusts their strategy in

response to Apple's entry, their Play Store version might also change, contaminating the control group. However, platform-specific development teams, different technology stacks, and distinct user bases often lead to independent strategies across platforms. The analysis I present here includes all apps regardless of multi-homing status, maximizing sample size while acknowledging this potential limitation.[13]

While the concerns so far bear on the validity of the control group, a third bears on the estimator. I estimate two-way fixed effects (TWFE) models on each margin, and with staggered timing and heterogeneous treatment effects, TWFE estimators can place negative weight on comparisons that use already-treated units as controls—the "forbidden comparison" problem (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). The never-treated Play Store control group limits the scope for such comparisons in every specification, and the incumbent specification rules them out entirely, comparing App Store apps only to that never-treated group within the same market and month. I state each case where the specifications are introduced (see Sections 3.1.1 and 3.1.3) and report the heterogeneity-robust estimators of Sun and Abraham (2021) and Wooldridge (2025), which avoid already-treated controls by construction, alongside TWFE for the incumbent and entry/exit margins, with pooled estimates collected in Appendix D.2.

A fourth concern is anticipation. Apple typically announces major operating-system features at its Worldwide Developers Conference in June and ships them at the September iOS release, so developers may learn of an entry months before it occurs. In the population studied here, 18 of the 22 markets were publicly previewed by Apple before release, with a median lead of three months, so the concern applies broadly. In Appendix D.6, I re-estimate the primary specifications with treatment re-anchored to the earliest public announcement date and, separately, with the months immediately preceding entry excluded from estimation; the estimates align with the primary results under both approaches.

A final identification concern is that Apple chooses where and when to enter. Its entry locations are not random draws from the App Store. Halckenhaeusser et al. (2026) find that Apple's iOS entries target categories characterized by low user satisfaction, limited innovation, and high market concentration. The relevant question, however, is what Apple selects on. Selection on market characteristics—entering prominent markets, or markets with weak incumbents—poses no threat: every comparison is between App Store and Play Store apps in the same market, so any feature of the markets Apple chooses is common to treated and control units by construction. Selection on market trajectories is likewise benign so long as the trajectory belongs to the market rather than to one platform, since a market-wide trend moves treated and control apps together and is absorbed in every specification I estimate. What the design does not protect against is selection on *platform-specific* trajectories—Apple targeting markets where the App Store side specifically is deteriorating relative to the Play Store side—which would be observationally equivalent to a treatment effect. The event studies reported throughout the paper display this differential component directly in their

[13] In Appendix E, I restrict the sample to single-homing apps (those available on only one platform) and confirm that multi-homing spillovers do not drive the main findings.

Table 3: Entry and Exit

	(1) Log Entry Count	(2) Log Exit Count
ATT	-0.2487***	0.0167
	(0.0756)	(0.0718)
Baseline Mean	1.499	1.250
N	792	792

Notes: Two-way fixed effects (TWFE) DiD estimates at the market-platform-month level. Standard errors clustered at the market-platform level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

pre-entry leads, and the anticipation checks above guard against the residual concern that pre-entry movement in those leads reflects early responses to an announced entry rather than selection.

3.1 Average Effects of Platform-Owner Entry

I begin by documenting average treatment effects for each set of outcomes: entry and exit dynamics, the characteristics of entering cohorts, and incumbent app responses.

3.1.1 Entry and Exit

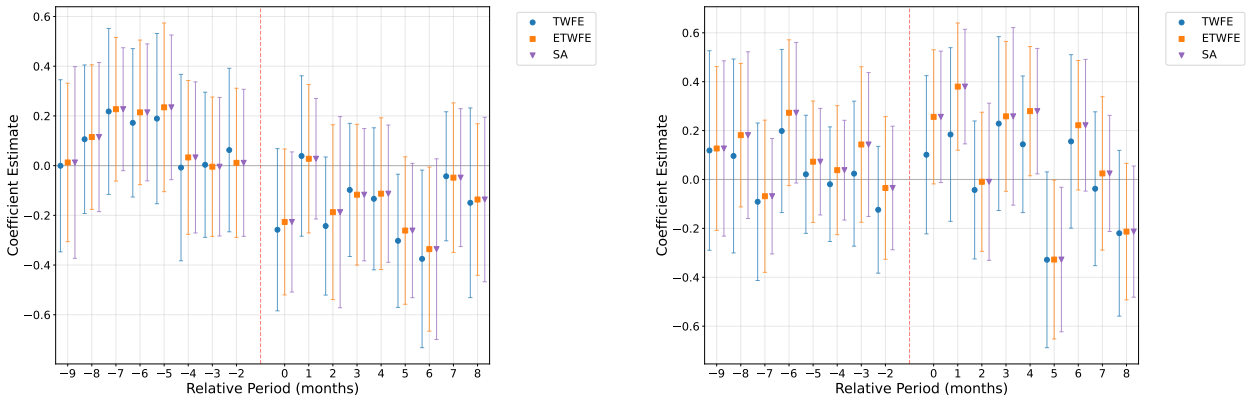
I first examine how Apple's entry affects market structure through changes in the number of third-party apps entering and exiting markets. To do so, I estimate the two-way fixed effects (TWFE) model:

$$y_{apt} = \alpha \cdot D_{apt} + \delta_{ap} + \eta_t + \varepsilon_{apt}, \tag{2}$$

where y_{apt} is the outcome (log entry or exit count) for market a on platform p at month t. The treatment indicator $D_{apt} = \mathbf{1}[p = \text{App Store}, t \geq T_a]$ equals one for App Store observations in periods at or after market a's entry date T_a, and zero otherwise; Play Store observations have $D_{apt} = 0$ throughout, serving as a never-treated control group. δ_{ap} are market-platform fixed effects, η_t are year-month fixed effects, and α is the coefficient of interest. The outcome variables are market-level counts of new app entries and app exits, transformed using $\ln(x+1)$ to handle zeros. Standard errors are clustered at the market-platform level. I estimate Equation (2) on a balanced market-platform-month panel dataset constructed from the market dynamics sample covering the 9 months before and after each entry event, using OLS.[14]

[14]This specification enters δ_{ap} and η_t separately rather than interacted, so an early-entering market's post-period observations can, in principle, help identify a later-entering market's effect through the shared calendar-month term—the mechanism behind the "forbidden comparison" problem in staggered designs (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). The never-treated Play Store group limits this risk, since a clean cross-platform comparison is always available, but does not eliminate it. I therefore do not rest the case on TWFE alone: I also estimate these effects using the heterogeneity-robust estimators of Wooldridge (2025) and Sun and Abraham (2021), presenting event study estimates in the figures below and estimates of the average effects in Appendix D.2.

Figure 4: Entry and Exit Dynamics



(a) Log Entry Count

(b) Log Exit Count

Note: Event study estimates showing dynamic treatment effects relative to one month before entry. TWFE is the primary estimator; ETWFE (Wooldridge, 2025) and SA (Sun and Abraham, 2021) shown for comparison. Analysis uses the market dynamics sample at the market-platform-month level. Standard errors are clustered at the market-platform level. Error bars represent 95% confidence intervals.

Table 3 presents the estimates, and Figure 4 presents event study estimates for entry and exit. I find a large and statistically significant decline in third-party entry following Apple's entry. There is a roughly 22% reduction in new app entry relative to the pre-entry baseline. This indicates that Apple's presence substantially deters new development in affected markets.[15] The exit regression, by contrast, shows no evidence of a change in exit rates. The absence of an exit effect, combined with the strong deterrence of entry, suggests that while Apple's entry discourages new competitors from entering the market, incumbent apps remain—potentially due to sunk development costs, the low cost of maintaining a listing, or the strategic value of maintaining platform presence. In Section 3.1.3, I consider an alternative measure of developers' participation and effort on the platform, the rate of product updating.

3.1.2 Composition of Entering Cohorts

Platform-owner entry can affect extensive-margin behavior by more than just impacting entry and exit counts—it may alter the composition of new entrants. I examine how entering cohorts differ before versus after Apple's entry using a cross-sectional difference-in-differences design. Each entering app is observed once, at entry. I estimate the model

$$y_j = \beta \cdot P_j + \alpha \cdot D_j + \delta_{ap} + \eta_t + \varepsilon_j, \tag{3}$$

using OLS. y_j represents entry characteristics for app j. $P_j = \mathbf{1}[t_j \geq T_{a(j)}]$ indicates whether app j entered at or after Apple's entry, regardless of platform; β captures the common post-entry shift.

[15] In Appendix D.6, I examine the pre-entry dynamics behind this estimate and show that it reflects a post-entry level shift rather than the continuation of platform-specific pre-entry trends.

The treatment indicator $D_j = \mathbf{1}[j \in \text{App Store}, t_j \geq T_{a(j)}]$ equals one for App Store apps that entered at or after their market's treatment date, and zero otherwise, where t_j is app j's entry date and $a(j)$ denotes its market. Play Store apps have $D_j = 0$ throughout. δ_{ap} are market×platform fixed effects and η_t are calendar month fixed effects—the same structure, and the same caveat, as in Section 3.1.1. The coefficient α captures the differential change in entry characteristics for App Store apps relative to Play Store apps following Apple's entry. Standard errors are clustered at the market-platform level.

The analyses that follow examine treatment effects along two broad dimensions. Monetization outcomes include the download price (in dollars, including zero-price apps), whether an app is offered for free, and whether it offers in-app purchases. Quality outcomes include the average user rating (1–5 stars), developer update frequency, and measures of consumer engagement: for entering cohorts, the initial level of rating counts and similarity to Apple's entrant; for incumbents, rating count growth (log-differenced).

Table 4 presents the results from estimating the cohort entry characteristics model Equation (3) using the market dynamics sample ($N = 4,262$ entering apps).[16] Entering cohorts adjust their monetization strategy following Apple's entry, with prices among new App Store entrants rising by about $0.25. This is likely driven, at least in part, by a decrease in the probability of offering a free app, which falls by an imprecisely estimated 5.8 percentage points, but is consistent with stronger evidence of a similar adjustment by incumbent apps (see Section 3.1.3). Importantly, though, these pooled effects average over markets that respond very differently. I formalize that cross-market variation in Section 3.2, and in Section 3.3 I show that the entering-cohort price response, like the incumbent one, is concentrated in markets Apple enters through OS integration. In-app purchase adoption shows no significant change.

Additionally, the average rating of new App Store entrants rises by about 0.30 stars relative to Play Store entrants following Apple's entry, among entrants with an observed first-month rating. Rating-count effects are positive but not statistically significant. The table also reports effects on new entrants' similarity (measured via cosine similarity) to the Apple entrant, indicating whether developers' strategic product space positioning decisions are affected by Apple's entry. I return to this seemingly null similarity effect in Section 3.2.

3.1.3 Incumbent Responses

I next examine how incumbent apps respond to Apple's entry across monetization and quality dimensions. All analyses below use the incumbent sample, defined in Section 2.4: a balanced monthly panel of apps operating on the platforms for the 9 months before and after Apple's entry. Because this balanced panel conditions on survival throughout the 18-month window, the incumbent ATTs estimate effects for apps that remain active through the post-entry period. In my primary

[16] I present event study estimates in Appendix B.

Table 4: Cohort ATT Estimates

	(1) Price	(2) Price (Paid Apps)	(3) Free App	(4) In-App Purchases	(5) Avg. Rating	(6) Log Rating Count	(7) Similarity
ATT	0.2544**	0.4317	-0.0578	-0.0292	0.2978**	0.1613	0.0016
	(0.1261)	(0.5812)	(0.0396)	(0.0373)	(0.1409)	(0.1326)	(0.0051)
Baseline Mean	0.3724	3.152	0.8818	0.1520	4.332	0.7963	0.6484
N	4,262	701	4,262	2,835	1,307	4,262	4,262

Notes: Cross-sectional DiD estimates for entering cohorts. Each app observed once at entry. Standard errors clustered at market-platform level shown in parentheses. For the IAP outcome, 5 markets are excluded due to insufficient Play Store in-app purchase data in the pre-period. The Rating column has fewer observations because some entrants receive no ratings in their first month. *** p<0.01, ** p<0.05, * p<0.10.

analysis, I estimate the TWFE model

$$y_{jt} = \alpha \cdot D_{jt} + \gamma_j + \eta_{a(j),t} + \varepsilon_{jt}, \tag{4}$$

using OLS.[17] y_{jt} is the outcome for app j in month t. The treatment indicator $D_{jt} = \mathbf{1}[j \in \text{App Store}, t \geq T_{a(j)}]$ equals one for App Store apps in periods at or after their market's entry date, and zero otherwise, where $a(j)$ denotes the market containing app j. Play Store apps have $D_{jt} = 0$ throughout, serving as a never-treated control group. γ_j are app fixed effects and $\eta_{a(j),t}$ are market-month fixed effects, which absorb market-specific time shocks.[18] This within-market structure—each App Store app is compared to never-treated Play Store apps in the same market and month—means no already-treated unit ever serves as a control, keeping this specification clear of the "forbidden comparison" problem that can affect staggered designs. The coefficient α represents the average treatment effect on the treated (ATT). Standard errors are clustered at the app level.[19]

I estimate effects on the same monetization and quality outcomes described above, adapted for the panel setting. For rating count, I use the log-difference transformation: $\Delta \ln(\text{Rating Count}+1)$, which captures the growth rate of cumulative ratings. Since new ratings arrive as users download and engage with apps, rating count growth serves as a proxy for consumer demand and ongoing engagement with the product.[20]

[17] See Appendix D.2 for results from alternative, staggered-treatment robust DiD estimators. Results are similar across methods.

[18] For robustness, I re-estimate the model using app and month (as opposed to market-month) fixed effects in Appendix D.3.

[19] I cluster at the app level to address within-app serial correlation. Because the analysis covers the full population of apps on both platforms, the design-based case for coarser clustering is weaker than in a sampled setting (Abadie et al., 2023). In Appendix D.4, I report cluster-jackknife (CRV3J) standard errors at the market-platform level instead. Quality results are robust to this alternative; monetization results retain their signs and magnitudes but are less precisely estimated, consistent with the concentration of pricing effects in a subset of markets I document in Section 3.2.

[20] I use rating count growth because no reliable app-level measure of downloads or revenue is available: the AppMonsta data used here records apps' chart ranks but not the underlying download or revenue counts, Apple and Google do not disclose those counts, and the download and revenue figures sold by other commercial vendors are extrapolated rather than directly observed.

Table 5: Incumbent App Outcomes

	(1) Price	(2) Price (Paid Apps)	(3) Free App	(4) In-App Purchases	(5) Avg. Rating	(6) Update	(7) Δ Log Rating Count
ATT	0.4221***	0.0842*	-0.0580***	-0.0061	0.0293***	0.0104	0.0170***
	(0.0464)	(0.0494)	(0.0092)	(0.0055)	(0.0069)	(0.0077)	(0.0035)
Baseline Mean	0.7484	4.090	0.8170	0.3704	3.978	0.2475	0.0470
N	59,616	16,187	59,616	46,260	56,770	59,616	59,616

Notes: Estimates from the incumbent analysis model at the app-month level. For the IAP outcome, 5 markets are excluded due to insufficient Android in-app purchase data in the pre-period. Standard errors clustered at the app level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 5 presents estimates for all incumbent outcomes.[21] Prices increase by an average of $0.42 following entry (column 1). Similar to what I found for the entering cohorts, this aggregate price increase masks important compositional changes. Apps become approximately 5.8 percentage points less likely to be free (column 3), a 7.1% decline. Conditional on charging a positive price, the price effect is modest (column 2), suggesting the overall price increase is driven primarily by apps switching from free to paid rather than existing paid apps raising prices. In-app purchase (IAP) adoption is essentially flat (column 4).[22] Taken together, the monetization estimates tell a consistent story: incumbents respond to Apple's entry by charging more, and they do so primarily on the extensive margin—switching from free to paid—rather than by raising prices among apps that already charge or by adding in-app purchases.

Columns 5–7 of Table 5 present estimates for the quality outcomes. Average ratings improve slightly (column 5), and update frequency is effectively unchanged (column 6). Rating count growth also shows a positive effect: an increase of approximately 0.017 on the log-difference scale, corresponding to roughly 1.7% additional growth (column 7). This suggests that Apple's entry may expand overall category visibility or demand, benefiting both incumbents and the platform.

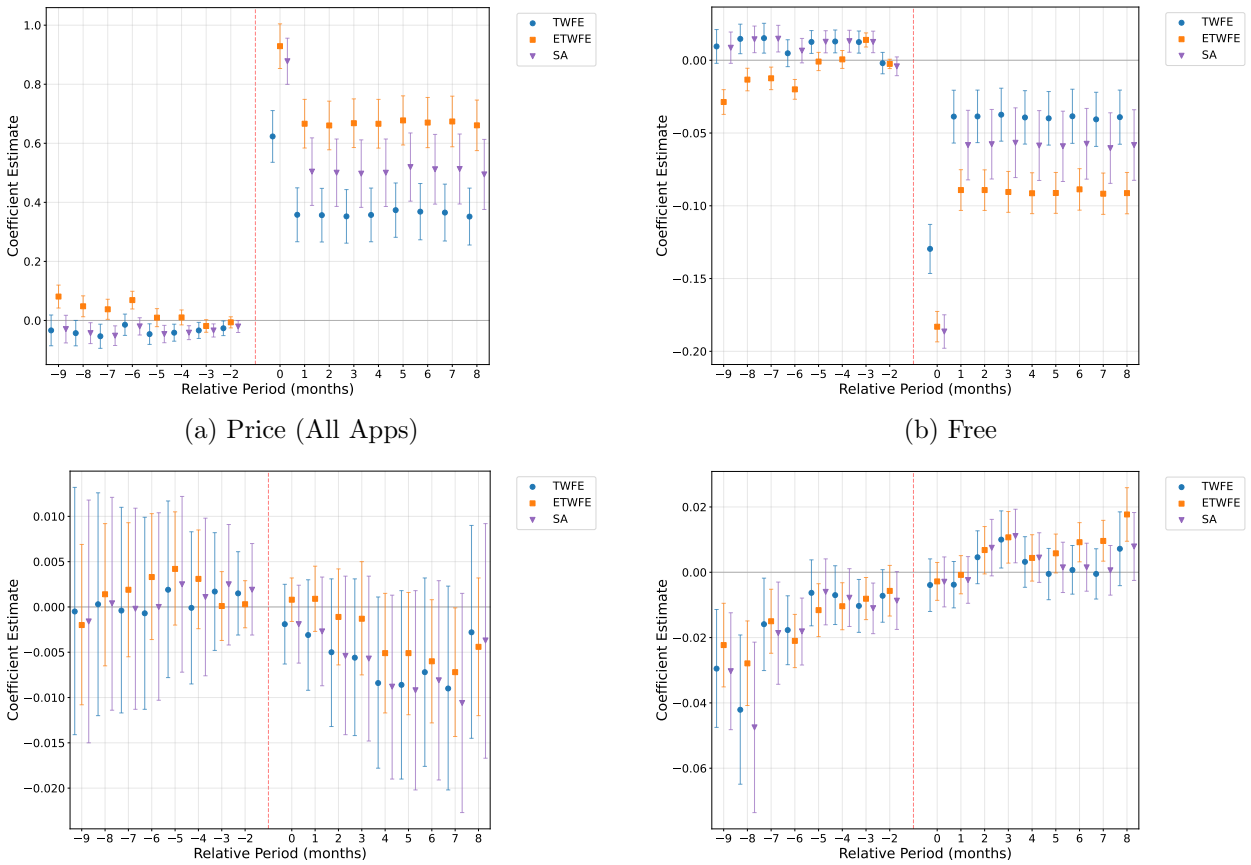
Figure 5 presents event studies for the principal incumbent outcomes; Figure D.1 reports the remaining outcomes (paid-app price, average rating, and update frequency). Estimates are consistent across estimators. The price increase and the decline in the proportion of free apps both show large initial responses that settle at less extreme but still substantial levels. This suggests developers may engage in some experimentation immediately following Apple's entry into their market. IAP adoption shows a modest, imprecisely estimated decline that flattens around the fourth post-treatment month.

The direction of this pricing response is itself notable. The textbook intuition is that the entry of a new, well-resourced competitor—particularly one that gives its product away for free—should push prices down. The opposite occurs, and several (not mutually exclusive) mechanisms can account for it. First, Apple's entry may sort consumers rather than simply intensify competition.

[21] These estimates use the full incumbent sample including multi-homing apps. In Appendix E, I show that restricting to single-homing apps yields similar results, confirming that cross-platform spillovers do not drive the findings.

[22] The AppMonsta data does not include prices for in-app purchases on the Google Play Store, only indicators of their presence. I am therefore unable to estimate the effect of Apple's entry on in-app purchase prices within this difference-in-differences framework.

Figure 5: Incumbent Treatment Effect Dynamics



(a) Price (All Apps)

(b) Free

(c) In-App Purchases

(d) Rating Count Growth

Note: Event study estimates from the incumbent analysis model, Equation (4). TWFE and SA include app and market×month fixed effects. ETWFE (Wooldridge, 2025) uses app and month fixed effects; its saturated market×time treatment interactions already absorb market-specific time variation. SA (Sun and Abraham, 2021) shown for comparison. Analysis is at the app-month level. For IAP, 5 markets are excluded due to insufficient Play Store in-app purchase data in the pre-period. Standard errors are clustered at the app level. Error bars represent 95% confidence intervals. Event studies for the remaining outcomes appear in Figure D.1.

Because Apple's offering is free, it is the closest substitute for the most price-sensitive users in a market, so as Apple absorbs that tier, the residual demand facing surviving incumbents is smaller but less price-sensitive, and the profit-maximizing response is to charge for access to it. Indeed, the notion that entry can raise rather than lower equilibrium prices when it reshapes the composition of demand dates back to Rosenthal (1980) and, more recently, Chen and Riordan (2008). The pattern has empirical precedent as well: Ward et al. (2002) find that name-brand prices did not fall but *rose* as private-label goods spread across grocery shelves alongside the national brands a retailer carries—a brick-and-mortar antecedent of the entry I study—and Seamans and Zhu (2014) find that the local newspapers most exposed to Craigslist's free classifieds raised subscription prices and differentiated their content rather than competing on price. This sorting also explains why the increase runs almost entirely through apps moving from free to paid rather than through paid apps raising prices: the apps with the most to gain from charging are those that had been giving the product away to the marginal user Apple now serves.

Second, the response may be dynamic. Before entry, a developer has reason to keep its price low, even zero, to build an installed base it expects to monetize later; Apple's entry forecloses that growth—the users a developer might have attracted are the ones Apple now captures for free—and shifts it from investing in its base toward harvesting it (Klemperer, 1987; Cabral, 2016). This reading fits the joint pattern in Section 3.1.1, where new entry falls sharply but exit does not: the return to investing in a new base falls, while the existing, partially captive base remains worth monetizing. Finally, developers may be repositioning—differentiating away from a free, mass-market first-party product, upward in price and quality, into a more specialized paid niche (Shaked and Sutton, 1982; Johnson and Myatt, 2003), consistent with the simultaneous, if modest, improvement in ratings.

What this response means for consumers is genuinely ambiguous, and the empirical approach I take here cannot adjudicate it. A shift from free to paid can reflect the exploitation of a captive segment—higher prices borne by the inframarginal users who remain after Apple skims away their more price-sensitive neighbors—or a welfare-improving move away from advertising- and data-supported monetization toward direct payment, bundled with the quality improvements documented above. Separating these readings would require demand-side evidence on willingness to pay, advertising load, or data practices that I do not observe, and is beyond the scope of this paper. I would also caution that the positive demand response is consistent with a more benign account in which Apple's entry validates the category and raises willingness to pay broadly, with no role for lock-in at all. The distance results in Section 3.4 bear on this question without fully resolving it: the move from free to paid is broadly distributed across competitive distance, as a market-wide sorting story would predict, while the gains in demand and ratings accrue to apps that remain differentiated from Apple's product, with the closest head-to-head competitors, if anything, harmed.

These pooled estimates provide useful benchmarks, but they may conceal substantial cross-market variation. Understanding whether the averages mask opposing responses across markets—

and what drives any such differences—has direct implications for the design of platform regulation. I turn to this question next.

3.2 Cross-Market Heterogeneity

The average effects documented thus far need not describe any individual market's experience. If markets differ in pre-existing competitive conditions, the nature of Apple's product, or the strength of consumer demand, the effect of entry may vary substantially—and markets may even respond in opposing directions. To quantify cross-market heterogeneity in treatment effects, I generalize Equations (2) to (4) by allowing market-specific ATTs $\alpha_{a(j)}$, interacting the treatment indicator with market dummies.[23]

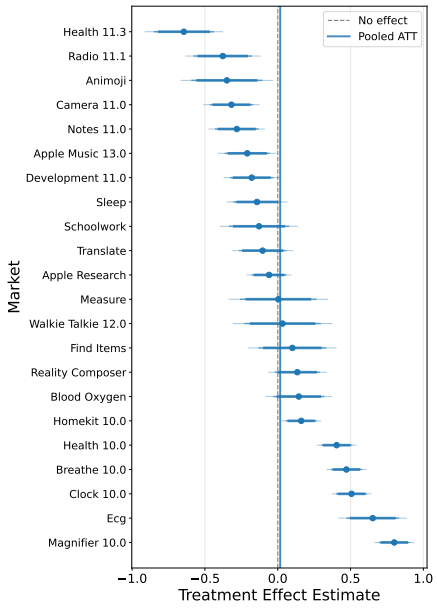
To characterize cross-market variation in treatment effects, I summarize the market-specific ATTs in two ways. First, I report three dispersion diagnostics commonly used in meta-analysis, applied to the inverse-variance weighted set of market-specific estimates. Cochran's Q tests the null that all markets share a common effect; under that null it follows a χ^2 distribution with $A-1$ degrees of freedom, where A is the number of markets, and I report its p-value. The I^2 statistic gives the share of total cross-market variation—from 0 to 100%—attributable to genuine between-market heterogeneity rather than sampling error; it is floored at zero, so outcomes with no detectable heterogeneity read $I^2 = 0$ (Higgins and Thompson, 2002; Higgins et al., 2003). The DerSimonian–Laird estimate of τ^2 (DerSimonian and Laird, 1986) is the variance of true effects across markets; I report $\sqrt{\tau^2}$, the between-market standard deviation, because it is directly comparable in scale to the pooled ATTs reported in the preceding tables. Second, I classify the markets behind each pooled effect into three groups: those whose market-specific effect is statistically significant at the 5% level and shares the sign of the pooled effect (% Confirming), those significant with the opposite sign (% Opposing), and those not statistically distinguishable from zero (% Null). Because a market with a precisely estimated null effect neither confirms nor contradicts the pooled sign, this partition keeps such markets separate rather than counting them as agreement, as a simple share of matching point estimates would.

Figure 6 presents market-specific treatment effects for four illustrative outcomes.[24] The exit result is perhaps the most striking: the near-zero effect documented in Section 3.1.1 masks genuinely opposing dynamics across markets. Among the markets where the exit effect is statistically distinguishable from zero, more show a significant decrease than a significant increase, with 32% opposing and 27% confirming the small positive pooled effect, so the null average reflects markets pulling in opposite directions rather than a common muted response. Similarity to the Apple entrant, for which the pooled cohort effect was essentially zero, shows a related pattern: most markets show no significant repositioning, while among those that do, more see new entrants cluster toward

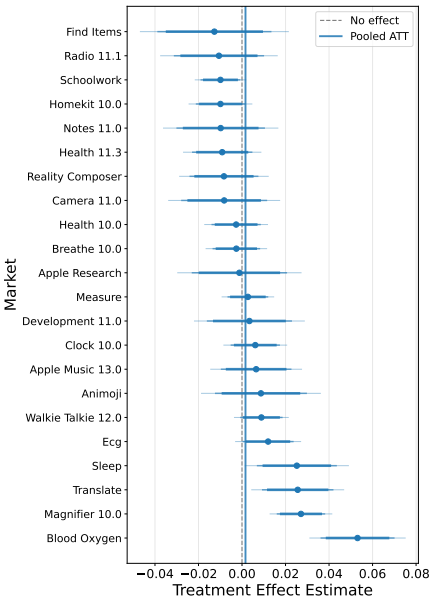
[23] A market with no within-market variation in a given outcome has no estimable market-specific effect for it—the few markets composed entirely of free apps, for instance, have no variation in price or in free/paid status. These markets are dropped from that outcome's analyses, which are computed over the markets in which the outcome varies.

[24] See Appendix C for the remaining figures.

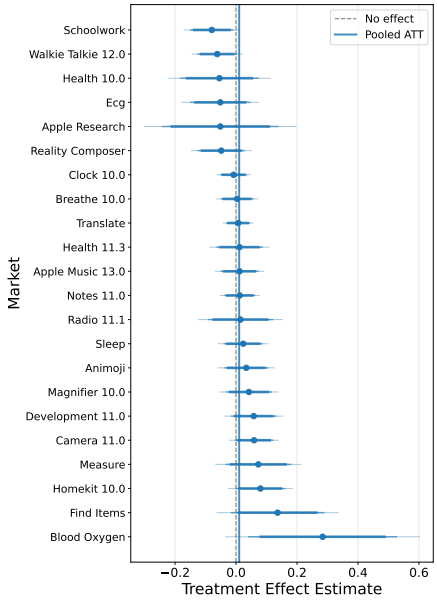
Figure 6: Selected Market-Specific Treatment Effects



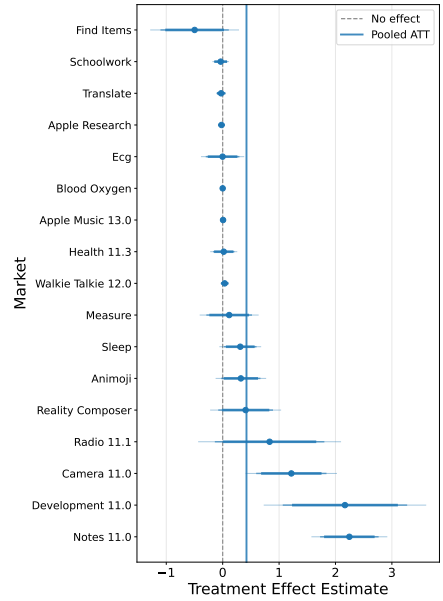
(a) Exit Count



(b) Similarity to Apple Entrant



(c) Incumbent Update Frequency



(d) Incumbent Price

Note: Market-specific treatment effects from Equations (2) to (4). Each panel plots market-level estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines); the solid vertical line marks the pooled ATT and the dashed line marks zero. Panel (a) uses the market dynamics sample and panel (b) the entering-cohort sample, both clustered at the market-platform level; panels (c) and (d) use the incumbent sample, clustered at the app level. Forest plots for all outcomes appear in Appendix C.

Apple's position than move away. By contrast, incumbent update frequency reflects a genuine near-absence of response: 86% of markets are statistically indistinguishable from zero, consistent with a modest, relatively uniform effect that is difficult to detect with market-level precision. The incumbent price response, by contrast, is highly concentrated: a few markets account for most of the pooled increase while the majority cluster near zero, so the \$0.42 average overstates the effect in the typical market.

Table 6 reports the full set of meta-analytic diagnostics across all three margins. The pattern is pervasive: most outcomes exhibit substantial cross-market heterogeneity, indicating that the pooled ATTs from Section 3.1 are consistently poor summaries of any individual market's experience. Following the taxonomy of Higgins et al. (2003), monetization heterogeneity is high for the unconditional price and free/paid margins ($I^2 \geq 81\%$ across both the cohort and incumbent analyses) and more moderate for incumbent in-app purchases ($I^2 = 53\%$, though the Q-test still rejects homogeneity there). Quality heterogeneity is more contained, with the incumbent quality outcomes at I^2 between 37% and 40% and cohort rating very high at $I^2 = 94\%$; entry and exit effects show substantial heterogeneity, with exit displaying $I^2 = 96\%$. The Q-test rejects homogeneity for every outcome except for the conditional price among paid incumbents.

The cohort similarity result deserves particular attention. The pooled effect is null: 73% of markets show no repositioning distinguishable from zero, so the null owes more to markets that do not move than to markets that offset one another. Among those that do respond, 23% cluster toward Apple's position and 5% move away; the high $I^2 = 79\%$ confirms that this split is genuine rather than sampling error. One interpretation is that Apple's entry may serve a coordinating function in some markets—signaling consumer demand, validating product categories, or concentrating developer effort around capabilities Apple has demonstrated or wishes to improve on its platform. In these markets, new entrants cluster toward Apple because its entry reduces uncertainty about viable product directions. In other markets, Apple's presence may instead push developers to seek new niches, with new entrants differentiating away from the now-occupied product position. The heterogeneity in similarity effects across markets is consistent with both mechanisms operating in different contexts, a finding that would be obscured by relying solely on the pooled (null) effect.

Finally, the conditional price among paid incumbents is the only effect in the table that both diagnostics call homogeneous. The pooled effect is itself modest (about \$0.08). That this intensive-margin price effect is both modest and undifferentiated across markets is consistent with the reading in Section 3.1.3 that the incumbent price response operates through the free-to-paid margin rather than through paid apps raising prices.

Having documented how widely Apple's effects vary across markets, I now turn to interpreting that variation along two dimensions the design lets me estimate directly. Section 3.3 shows that how Apple enters—as a standalone app or an OS integration—shapes the magnitude of the monetization response for incumbents and new entrants alike, and Section 3.4 shows that an app's competitive proximity to Apple's offering shapes how strongly it is affected. While these are just two of many dimensions along which markets can differ, they are dimensions along which platform entry's ef-

Table 6: Cross-Market Heterogeneity Diagnostics

Outcome	Est. (SE)	% Conf.	% Opp.	% Null	I^2 (Q p)	$\sqrt{\tau^2}$
Panel A: Entry and Exit						
Entry Count	-0.249 (0.076)	64%	9%	27%	88% (<0.001)	0.290
Exit Count	0.017 (0.072)	27%	32%	41%	96% (<0.001)	0.373
Panel B: Entering Cohorts						
Monetization						
Price	0.254 (0.126)	32%	9%	59%	81% (<0.001)	0.009
Free	-0.058 (0.040)	32%	14%	55%	83% (<0.001)	0.004
In-App Purchases	-0.029 (0.037)	24%	29%	47%	79% (<0.001)	0.067
Price (Paid)	0.432 (0.581)	29%	21%	50%	90% (<0.001)	1.879
Quality						
Rating	0.298 (0.141)	40%	10%	50%	94% (<0.001)	0.441
Rating Count	0.161 (0.133)	27%	14%	59%	81% (<0.001)	0.374
Similarity	0.002 (0.005)	23%	5%	73%	79% (<0.001)	0.013
Panel C: Incumbent Responses						
Monetization						
Price	0.422 (0.046)	25%	0%	75%	88% (<0.001)	0.127
Free	-0.058 (0.009)	23%	0%	77%	82% (<0.001)	0.022
In-App Purchases	-0.006 (0.005)	0%	19%	81%	53% (0.007)	0.009
Price (Paid)	0.084 (0.049)	0%	0%	100%	0% (0.504)	0.000
Quality						
Rating	0.029 (0.007)	14%	5%	82%	38% (0.037)	0.021
Update Freq.	0.010 (0.008)	9%	5%	86%	40% (0.027)	0.028
Rating Count Growth	0.017 (0.004)	23%	0%	77%	37% (0.040)	0.011

Notes: This table reports heterogeneity diagnostics applied to market-level ATTs from TWFE regressions. Panel A uses the market dynamics sample at the market-platform-month level. Panel B uses the market dynamics sample at the app level (each app observed once at entry). Panel C uses the incumbent sample at the app-month level with market×month fixed effects. Est. is the pooled TWFE ATT with standard error in parentheses. For each outcome, markets are classified by the significance and direction of their market-specific effect: % Conf. is the share of markets significant at the 5% level with the same sign as the pooled estimate, % Opp. the share significant with the opposite sign, and % Null the share not significantly different from zero. Figure 6 reports 90%, 95%, and 99% confidence intervals for the market-level estimates. I^2 measures the proportion of total variation due to between-market heterogeneity rather than sampling error; Q p-value tests the null of homogeneous effects. $\sqrt{\tau^2}$ is the DerSimonian–Laird estimate of the between-market standard deviation of true effects.

fects vary systematically and to an economically meaningful degree—and both are observable to a regulator, making them natural margins on which targeted oversight could condition.

3.3 Entry Type and Competitive Dynamics

The cross-market heterogeneity documented in Section 3.2 motivates examining whether observable market characteristics explain the variation in treatment effects. I investigate one salient dimension: whether Apple enters via a new standalone app or OS integration. Entry type varies across markets and has clear theoretical implications for competitive dynamics.

To directly test whether entry type affects competitive outcomes among incumbent apps, I estimate a variation of Equation (4) with entry type interactions,

$$y_{jt} = \alpha_{\text{SA}} \cdot D_{jt} + \alpha_{\text{Diff}} \cdot D_{jt} \times \text{Integrated}_j + \gamma_j + \eta_{a(j),t} + \varepsilon_{jt}, \tag{5}$$

where D_{jt} is defined as in Equation (4) and Integrated_j indicates whether app j is in a market where Apple entered via OS integration rather than a standalone app. The coefficient α_{SA} captures the treatment effect for standalone entry (the baseline group), α_{Diff} captures the additional effect for integrated entry, and their sum $\alpha_{\text{SA}} + \alpha_{\text{Diff}}$ yields the total effect for integrated entry. Standard errors are clustered at the app level. This specification directly tests the hypothesis that entry type matters: a significant α_{Diff} indicates systematically different responses. I present the results in Table 7. The "All" column reproduces the pooled ATT from the baseline specification (Equation (4)) for reference, while the remaining columns report estimates from the interaction specification (Equation (5)).

Two key patterns emerge from Table 7. First, integrated entry generates larger price effects than standalone entry.[25] These differences suggest distinct competitive mechanisms: integrated entry forces more aggressive pricing responses, while standalone entry generates more modest adjustments. Second, the entry types exhibit similar demand responses, as proxied by the growth rate of rating counts. This suggests that any attention-grabbing demand expansion associated with Apple's entry into a submarket does not necessarily depend on how Apple implements entry.

The pattern is not confined to incumbents. Re-estimating the entry-type interaction on the entering-cohort sample, I find the same split: the entering-cohort price response is concentrated in integrated-entry markets, where entrants raise prices and shift away from free pricing, while entrants in standalone markets do not (Table 8).[26] The price difference between entry types is statistically significant and comparable in magnitude to the incumbent difference, while the free/paid difference among entrants points the same way but is less precisely estimated. That the same monetization

[25] The cross-entry-type comparison treats standalone and integrated markets symmetrically; Section 2.3 describes the synthesis procedure used to construct integrated-market description vectors and reports embedding-space validation that these vectors are comparable to standalone-entry App Store descriptions. Appendix D.5 further shows that this entry-type differential is not an artifact of the $\theta = 0.6$ market definition.

[26] The rating-count outcome necessarily differs across the two samples. For incumbents it is the monthly log-difference, a growth rate; for entrants, who have no prior month to difference against, it is the log level. The two are not directly comparable.

Table 7: Entry Type Effects on Incumbent Apps: Standalone vs. Integrated

Outcome	All	Integrated	Standalone	Difference
Price	0.422***	0.586***	0.049*	0.537***
	(0.0464)	(0.0677)	(0.0286)	(0.0735)
Free	-0.058***	-0.076***	-0.0013	-0.075***
	(0.0092)	(0.0135)	(0.0040)	(0.0141)
In-App Purchases	-0.0061	-0.0008	-0.011	0.010
	(0.0055)	(0.0042)	(0.0071)	(0.0083)
Price (Paid Apps)	0.084*	0.016	0.056	-0.040
	(0.0494)	(0.0494)	(0.0551)	(0.0740)
Update	0.010	0.024	-0.0002	0.024
	(0.0077)	(0.0177)	(0.0222)	(0.0283)
Rating	0.029***	0.0094	0.011	-0.0020
	(0.0069)	(0.0062)	(0.0084)	(0.0105)
Δ Log Rating Count	0.017***	0.0038	-0.0013	0.0051
	(0.0035)	(0.0045)	(0.0041)	(0.0061)

Notes: TWFE estimates with standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10. All: 22 markets (10 standalone + 12 integrated). Difference: Integrated − Standalone.

split recurs in this independent setting—present under integrated entry, muted or absent under standalone entry—shows the entry-type effect to be broad-based, not specific to incumbents: how Apple enters, and not only whether it enters, shapes the competitive response.

One possible explanation for the price differences is the unavoidability of integrated releases. OS-integrated features cannot be separately uninstalled or avoided by users, creating fundamentally different competitive pressure from that of marketplace apps. When Apple releases a standalone app like Translate, users must actively discover it in the App Store, evaluate it against alternatives, and choose to download it. Market forces can operate normally: if third-party alternatives are superior, users may never adopt Apple's offering.[27]

Integrated entry eliminates this choice architecture. When Apple integrates translation capabilities directly into iOS, every user automatically possesses this functionality upon updating their device. Third-party translation apps must now compete against a feature that users already have, cannot remove, and may discover through system prompts or Siri suggestions. This unavoidability likely explains the larger monetization responses to integrated entry: facing an unremovable competitor, third-party apps must differentiate more aggressively through pricing and business model

[27] Of course, self-preferencing by the platform, such as favorable search placement or default settings, may blunt this competitive discipline even for standalone entries.

Table 8: Entry Type Effects on Entering Cohorts: Standalone vs. Integrated

Outcome	All	Integrated	Standalone	Difference
Price	0.254**	0.414***	-0.138	0.551***
	(0.1261)	(0.1353)	(0.1290)	(0.1336)
Price (Paid Apps)	0.432	1.57**	0.126	1.45**
	(0.5812)	(0.6452)	(0.5501)	(0.5771)
Free	-0.058	-0.087**	0.013	-0.100
	(0.0396)	(0.0419)	(0.0600)	(0.0624)
In-App Purchases	-0.029	-0.018	-0.044	0.025
	(0.0373)	(0.0337)	(0.0518)	(0.0453)
Rating	0.298**	0.349**	0.169	0.180
	(0.1409)	(0.1763)	(0.1663)	(0.2223)
Log Rating Count	0.161	0.150	0.190	-0.040
	(0.1326)	(0.1406)	(0.1574)	(0.1313)
Similarity	0.0016	-0.0009	0.0078	-0.0087*
	(0.0051)	(0.0052)	(0.0061)	(0.0052)

Notes: TWFE estimates with standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10. All: 22 markets (10 standalone + 12 integrated). Difference: Integrated − Standalone.

changes.[28]

3.4 Product Space Dynamics

Notably, not all apps in an affected market are equally exposed to, or "treated" by, platform entry. An app providing nearly identical functionality to Apple's entrant may face different competitive pressures than a tangentially related app at the market periphery. I exploit the continuous similarity measure from the SBERT embeddings described in Section 2.3 to estimate how treatment effects vary with competitive proximity. I use B-splines to flexibly estimate this relationship within the difference-in-differences framework.

For each app j in market a with cosine similarity $d_j \in [\theta, 1]$ to the entrant, I estimate similarity heterogeneity by interacting K B-spline basis functions with the treatment indicator:

$$y_{jt} = \sum_{k=1}^{K} \phi_k \cdot D_{jt} \cdot B_k(d_j) + \gamma_j + \eta_{a(j),t} + \varepsilon_{jt}. \tag{6}$$

D_{jt} is defined as above, $B_k(\cdot)$ are B-spline basis functions, K is the number of basis functions (degrees of freedom), γ_j are app fixed effects, and $\eta_{a(j),t}$ are market×month fixed effects. The inclusion of market×month fixed effects ensures that distance-invariant market-wide shocks are absorbed, so that the treatment-effect profile $\tau(d)$, defined in Equation (7) below, captures variation along the similarity gradient net of market-wide trends. This specification assumes that, absent treatment, the iOS–Android outcome gap does not vary systematically with similarity to Apple's product within a market—an assumption supported by the pre-trends documented in the earlier event studies. This approach allows me to flexibly estimate how the treatment effect varies in an app's similarity to Apple's first-party product without imposing a specific functional form.

I select K via cross-validation, which yields $K = 3$.[29] In the analysis below, I plot the treatment effect at each similarity level,

$$\tau(d) = \sum_{k=1}^{K} \phi_k \, B_k(d). \tag{7}$$

This captures the treatment effect for apps at similarity d, showing how competitive pressure varies

[28] Whether a particular user encounters an integrated feature depends on that user—on whether they update their device and whether they choose to use the feature. What matters for a developer is not that every user does, but that it cannot tailor a single product to those who do and those who do not. It offers one app, which competes with the integrated feature so long as a meaningful share of its users have received it. That share is large. Apple reports that roughly three-quarters of recent devices run the latest iOS version within a year of release (Apple Inc., 2026), so a developer serving that base competes with the integrated feature regardless of any one user's choice to upgrade or to use it.

[29] Specifically, I use 10-fold cross-validation at the app level, testing values from 2 to 8 and applying the one-standard-error rule (Breiman et al., 1984; Hastie, Tibshirani, and Friedman, 2009): among all candidate values, I choose the most parsimonious K whose mean cross-validation error falls within one standard error of the minimum. This favors simpler models when additional complexity yields negligible predictive improvement, guarding against overfitting.

with distance from the entrant.[30]

Figure 7 presents $\tau(d)$ for all incumbent outcomes, showing how treatment effects vary with competitive proximity to the first-party entrant. The unconditional price effect (panel (a)) is positive and statistically significant across most of the similarity range, with a non-monotonic shape that peaks among apps with moderately high similarity to Apple's entrant before reversing sharply—though imprecisely—for the most similar apps. The probability of being free (panel (c)) shows a relatively uniform decline of 5–6 percentage points that varies little with similarity, attenuating toward zero at the highest similarity levels though with wider confidence intervals, suggesting this monetization shift is broadly distributed rather than concentrated among the closest competitors. Conditional on being paid (panel (b)), the price effect is small and generally statistically insignificant, confirming that the overall price increase reflects apps switching from free to paid rather than paid apps raising prices. The IAP effect (panel (d)) is modestly negative throughout, becoming significantly larger for the most similar apps, consistent with the closest competitors reducing in-app purchase offerings in response to entry.

The quality panels reveal a distinct pattern. Ratings (panel (e)) are significantly positive in the mid-similarity range but turn sharply negative for the apps most similar to Apple's entrant. This is consistent with the closest competitors being most harmed by first-party entry, while more peripheral apps may benefit from complementarities with Apple's entrant or from offering a sufficiently differentiated, possibly premium, version of the product. Update frequency (panel (f)) shows effects near zero and generally insignificant across the similarity range, consistent with the small, homogeneous cross-market effect documented above. Finally, rating count growth (panel (g)) is positive and significant for moderately similar apps ($d < 0.75$), declining through zero and turning negative for the most similar competitors. This suggests the positive demand shocks associated with platform-owner entry tend to accrue to apps that maintain sufficient differentiation from Apple's product.

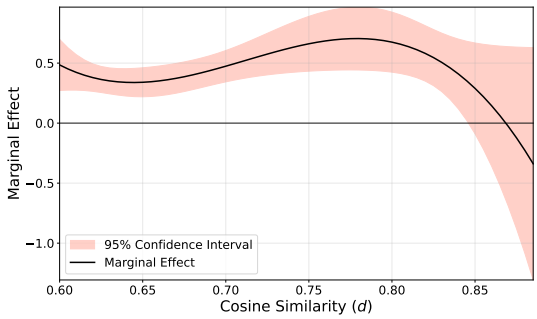
Taken together, the distance gradients for incumbents and the similarity heterogeneity for entering cohorts documented in Section 3.2 paint a consistent picture: Apple's entry reshapes the competitive landscape non-uniformly, with both the direction and magnitude of effects depending on a firm's proximity to the entrant and on market-specific conditions.

4 Concluding Discussion

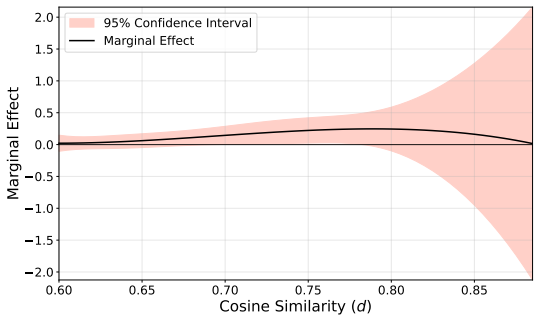
This paper provides systematic evidence on how third-party developers respond to platform-owner entry across multiple markets. Its central result is not any single average effect but the distribution of effects across markets. Studying 22 instances of Apple entering submarkets in its App Store—10 through standalone app releases and 12 through integrated OS features—I find that platform-owner entry deters new third-party entrants by 22% and reshapes incumbent monetization strategies, but

[30] I evaluate and plot $\tau(d)$ over the observed similarity range—from θ to the highest third-party similarity in the data—together with pointwise 95% confidence intervals (standard errors clustered at the app level). The curve is not extrapolated toward the entrant at $d = 1$, so its high-similarity end reflects the most similar *observed* competitors.

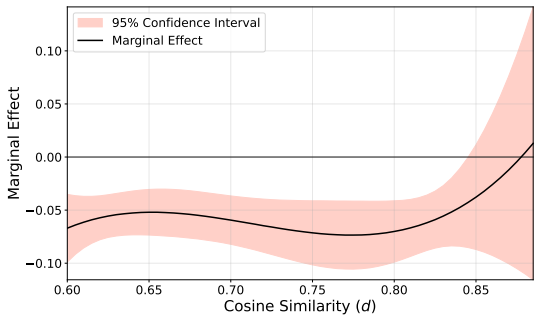
Figure 7: Incumbent Effects by Distance from Entrant



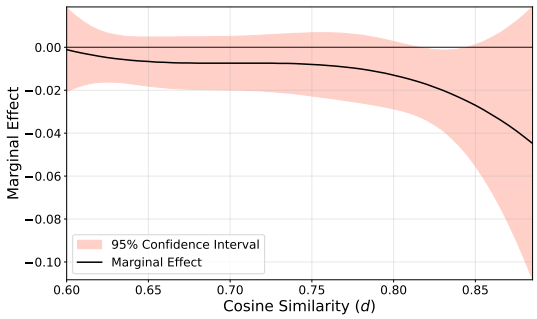
(a) Price (All Apps)



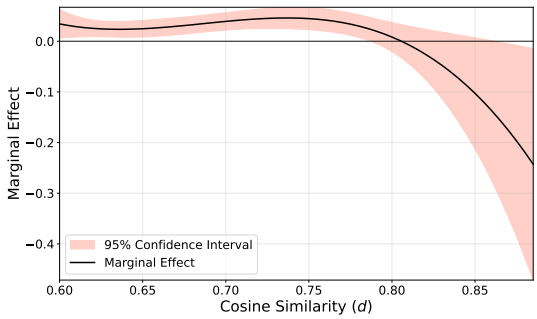
(b) Price (Paid Apps Only)



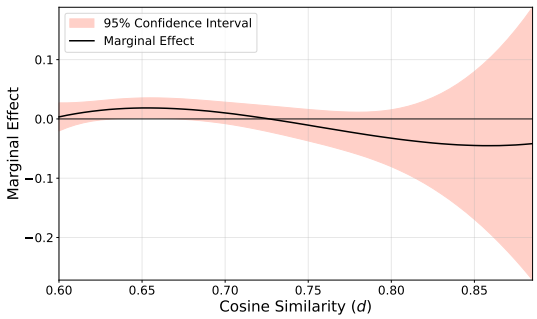
(c) Free



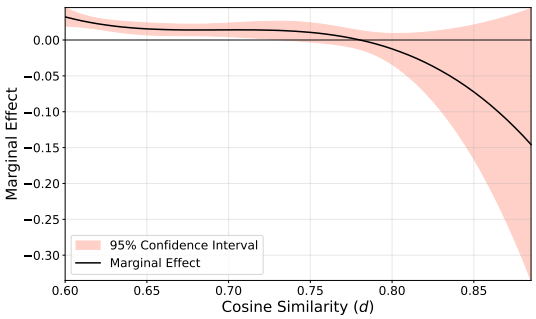
(d) In-App Purchases



(e) Rating



(f) Update Frequency



(g) Rating Count Growth

Note: Treatment effects from the incumbent distance model, Equation (6). For the IAP outcome, 5 markets are excluded due to insufficient Play Store in-app purchase data in the pre-period. Increasing cosine similarity indicates greater similarity to Apple's product, located at $d = 1$. Standard errors are clustered at the app level. Shaded areas represent 95% confidence intervals.

that these averages mask wide variation across markets and outcomes. The evidence reveals that platform entry is neither uniformly beneficial nor uniformly harmful, but rather generates widely varying effects that depend systematically on market context, competitive proximity, and entry implementation.

Cross-market heterogeneity is pervasive: meta-analytic diagnostics reveal high I^2 values across all three margins of analysis—entry/exit dynamics, cohort composition, and incumbent responses—indicating that average effects often poorly represent any individual market. Perhaps most surprisingly, I find that many markets experience no meaningful effects across several outcomes, with quality metrics particularly unaffected. This prevalence of null effects contradicts popular claims of either universal harm from foreclosure or overwhelming benefits from technology spillovers.

Within markets, competitive proximity shapes effects through distance gradients that themselves vary across markets. Entry type matters: integrated OS features generate pricing effects on the unconditional price and free/paid margins that are roughly an order of magnitude larger than those from standalone apps. New entrant positioning also reveals that Apple's entry reshapes the product space differently across markets, with entrants clustering toward or differentiating away from Apple's position depending on market context.

Several important limitations qualify these findings and their policy implications. Most fundamentally, the reduced-form approach employed here has both advantages and disadvantages. A key advantage is that it readily accommodates the wide range of (potentially conflicting) mechanisms playing out at once when a platform-owner enters its digital marketplace. This approach is not restricted to modeling just one or two of the relevant factors, instead providing insight into the net competitive effects of entry. However, this approach is also limited in that it cannot directly assess consumer welfare. Indeed, the combination and variation of effects documented here have ambiguous welfare implications. As I discuss for the incumbent pricing response in Section 3.1.3, the shift from free to paid could reflect either the exploitation of locked-in users or a welfare-improving move away from advertising- and data-funded models—a distinction reduced-form prices cannot settle. The absence of an average exit response, together with the repricing I find among incumbents that remain, is consistent with the differentiated nature of the market. Apple enters as an imperfect substitute, not an identical rival, so each incumbent holds a distinct position, with demand of its own.

The findings presented here open important avenues for future research. The stark heterogeneity I document—across markets, within markets by competitive distance, and between analysis margins—calls for deeper investigation into what determines these differential effects. The heterogeneous similarity results for new entrants documented in Section 3.2—where new entrants cluster toward or differentiate away from Apple's position depending on market context—suggest rich strategic interactions that other approaches could help disentangle. Understanding what drives the opposing exit responses across markets, or why similarity effects reverse sign, could help predict which markets are most vulnerable to platform competition and inform more targeted regulatory approaches. The systematic differences between standalone and integrated entry also warrant fur-

ther study, particularly as platforms increasingly blur the lines between operating system features and marketplace applications.

The longer-run dynamics of platform competition remain largely unexplored. The nine-month post-entry window considered here may miss important evolutionary patterns where initial defensive responses give way to accommodation, innovation, or exit. Technology spillovers from platform entry might take years to fully materialize, particularly when new APIs and frameworks require developers to learn new skills and rebuild applications. Understanding these dynamic patterns could reveal whether the immediate competitive harm from platform entry is offset by longer-run innovation benefits, or whether initial advantages compound over time into durable market dominance.

These findings provide crucial empirical evidence for the ongoing global debate over platform regulation and antitrust enforcement. The EU's Digital Markets Act restricts gatekeepers from self-preferencing (EUR-Lex, 2022), while the proposed American Innovation and Choice Online Act would similarly constrain platforms' ability to leverage nonpublic data against their business users (Congress.gov, 2022, 2026). The evidence reveals that calls for blanket restrictions (see, for example, Warren, 2019) fundamentally misunderstand the nature of platform competition. The prevalence of null effects demonstrates that categorical bans would block a great deal of entry for which there is no evidence of harm to developers in affected markets. More troublingly, the concentration of effects among close competitors means that broad statutory language could restrict entry whose effects reach only a narrow segment of apps, leaving most without a measurable response. Choi, Kim, and Mukherjee (2025) reinforce this caution against blanket rules, showing that an outright ban on a platform's use of third-party data can itself reduce consumer welfare, because the platform responds by raising the fees that govern sellers' entry, so the welfare consequences of such a ban cannot be signed without accounting for the platform's strategic response.

The differences between standalone and integrated entry provide actionable guidance for antitrust enforcement and regulatory design. Integrated OS features present fundamentally different competitive concerns than standalone apps that users can choose to ignore. This distinction maps directly onto antitrust theory: integrated features resemble tying-like arrangements or technological bundling that can increase foreclosure risk, while standalone apps are more likely to compete on merit in the marketplace.[31] Antitrust authorities should therefore apply different levels of scrutiny based on entry mode, with integrated features triggering heightened review similar to merger analysis in adjacent markets. The unavoidability of integrated features creates the kind of foreclosure concerns that antitrust law is designed to address, while standalone apps' market-based adoption provides the competitive constraints that normally obviate regulatory intervention.

The heterogeneity across markets further challenges the premise underlying many current regulatory proposals—that platform entry is categorically harmful to innovation and competition. Rather than categorical entry prohibitions, the findings support regulatory scrutiny calibrated to entry mode and competitive proximity—with integrated features warranting heightened attention

[31] Pre-installed standalone apps, however, may warrant greater scrutiny.

given their substantially larger effects, and market definition in platform cases accounting for the concentration of effects among close competitors.

Ultimately, as policymakers worldwide grapple with platform power, evidence-based approaches that recognize the heterogeneity in platform competition effects will be essential for crafting interventions that preserve innovation incentives while preventing anticompetitive harm. That heterogeneity carries a simple implication: a platform is not a monolith. It is a collection of many distinct submarkets that respond to entry in different, sometimes opposing, ways, and effective oversight has to work market by market, not platform-wide.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (1):1–35.

Anderson, Simon P. and Özlem Bedre-Defolie. 2024. "Hybrid platform model: monopolistic competition and a dominant firm." *The RAND Journal of Economics* 55 (4):684–718. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12478.

Apple Inc. 2026. "App Store." https://developer.apple.com/support/app-store/. Accessed 6 July 2026.

AppMonsta. 2016. "App Store Data and API." URL https://appmonsta.com. Accessed: 2024.

Arya, Anil, Brian Mittendorf, and David E. M. Sappington. 2007. "The Bright Side of Supplier Encroachment." *Marketing Science* 26 (5):651–659.

Bergès-Sennou, Fabian, Philippe Bontems, and Vincent Réquillart. 2004. "Economics of Private Labels: A Survey of Literature." *Journal of Agricultural & Food Industrial Organization* 2 (1).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1):249–275.

Boudreau, Kevin J. and Andrei Hagiu. 2009. "Platform Rules: Multi-Sided Platforms as Regulators." In *Platforms, Markets and Innovation*, edited by Annabelle Gawer. Edward Elgar Publishing, Inc., 163–191.

Breiman, Leo, Jerome Friedman, R. A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Cabral, Luís. 2016. "Dynamic Pricing in Customer Markets with Switching Costs." *Review of Economic Dynamics* 20:43–62.

Chen, Yongmin and Michael H. Riordan. 2008. "Price-Increasing Competition." *The RAND Journal of Economics* 39 (4):1042–1058.

Cheyre, Cristobal, Benjamin T. Leyden, Sagar Baviskar, and Alessandro Acquisti. 2026. "Did Apple's App Tracking Transparency Framework Harm the App Ecosystem?" *Working Paper* .

Choi, Jay Pil, Kyungmin Kim, and Arijit Mukherjee. 2025. ""Sherlocking" and Platform Information Policy." *Management Science* 71 (11):9812–9830.

Congress.gov. 2022. "Text – S. 2992 – 117th Congress (2021–2022): American Innovation and Choice Online Act." https://www.congress.gov/bill/117th-congress/senate-bill/2992/text. Accessed 23 June 2025.

———. 2026. "Text – S. 4746 – 119th Congress (2025–2026): American Innovation and Choice Online Act." https://www.congress.gov/bill/119th-congress/senate-bill/4746/text. Accessed 15 July 2026.

Crémer, Jacques, Yves-Alexandre de Montjoye, and Heike Schweitzer. 2019. "Competition policy for the digital era." Tech. rep., European Commission and Directorate-General for Competition. URL https://data.europa.eu/doi/10.2763/407537.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–2996.

De Chant, Tim. 2022. "Antitrust bill that bars Big Tech self-preferencing advances in Senate." *Ars Technica* URL https://arstechnica.com/tech-policy/2022/01/antitrust-bill-that-bars-big-tech-self-preferencing-advances-in-senate/.

DerSimonian, Rebecca and Nan Laird. 1986. "Meta-Analysis in Clinical Trials." *Controlled Clinical Trials* 7 (3):177–188.

Ershov, Daniel. 2024. "Variety-Based Congestion in Online Markets: Evidence from Mobile Apps." *American Economic Journal: Microeconomics* 16 (2):180–203. URL https://www.aeaweb.org/articles?id=10.1257/mic.20200347.

EUR-Lex. 2022. "Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)." http://data.europa.eu/eli/reg/2022/1925/oj. Official Journal L 265/1, 12 October 2022. Accessed 23 June 2025.

Förderer, Jens, Thomas Kude, Sunil Mithas, and Armin Heinzl. 2018. "Does Platform Owner's Entry Crowd Out Innovation? Evidence from Google Photos." *Information Systems Research* 29 (2):444–460.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2):254–277.

Gosalia, Anuj. 2018. "Announcing ARCore 1.0 and new updates to Google Lens." Google Developers Blog, https://developers.googleblog.com/2018/02/announcing-arcore-10-and-new-updates-to.html. Accessed 28 May 2026.

Gugler, Klaus, Florian Szücs, and Ulrich Wohak. 2024. "Using Natural Language Processing to Delineate Digital Markets." *Stanford Computational Antitrust* IV:33–52.

Gutt, Dominik, Jürgen Neumann, Wael Jabr, and Dennis Kundisch. 2019. "The App Updating Conundrum: Implications of Platform's Rating Resetting on Developers' Behavior." In *Proceedings of the Fortieth International Conference on Information Systems (ICIS)*. Munich, Germany.

Hagiu, Andrei, Tat-How Teh, and Julian Wright. 2022. "Should platforms be allowed to sell on their own marketplaces?" *The RAND Journal of Economics* 53 (2):297–327. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12408.

Halckenhaeusser, André, Jens Förderer, Armin Heinzl, and Ola Henfridsson. 2026. "Strategic Drivers of Core Expansion on Software Platforms: Evidence from Apple iOS." *Information Systems Research* 37 (2):1222–1241. URL https://doi.org/10.1287/isre.2023.0347.

Han, Sukjin, Eric Schulman, Kristen Grauman, and Santhosh Ramakrishnan. 2024. "Shapes as Product Differentiation: Neural Network Embedding in the Analysis of Markets for Fonts." Working paper.

Hansen, Bruce E. 2025. "Standard Errors for Difference-in-Difference Regression." *Journal of Applied Econometrics* 40 (3):291–309. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3110.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2nd ed.

Higgins, Julian P. T. and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21 (11):1539–1558.

Higgins, Julian P. T., Simon G. Thompson, Jonathan J. Deeks, and Douglas G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *BMJ* 327 (7414):557–560.

Hoberg, Gerard and Gordon Phillips. 2016. "Text-Based Network Industries and Endogenous Product Differentiation." *Journal of Political Economy* 124 (5):1423–1465.

———. 2025. "Scope, Scale, and Concentration: The 21st-Century Firm." *The Journal of Finance* 80 (1):415–466.

Jamison, Mark A. and Byoungmin Yu. 2026. "Choosing Product Space: Lessons from the App Economy." Working Paper.

Jiang, Baojun, Kinshuk Jerath, and Kannan Srinivasan. 2011. "Firm Strategies in the "Mid Tail" of Platform-Based Retailing." *Marketing Science* 30 (5):757–775.

Johnson, Garrett A., Tesary Lin, Liang Zhong, and James C. Cooper. 2026. "COPPAcalypse? The YouTube Settlement's Impact on Kids' Content." *Management Science* URL https://doi.org/10.1287/mnsc.2024.05295. Articles in Advance.

Johnson, Justin P. and David P. Myatt. 2003. "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning." *The American Economic Review* 93 (3):748–774.

Kapacinskaite, Aldona and Ahmadreza Mostajabi. 2024. "Competing with the platform: Complementor positioning and cross-platform response to entry." *Strategic Management Journal* 45 (12):2577–2607.

Kesler, Reinhold. 2023. "The Impact of Apple's App Tracking Transparency on App Monetization." *Working Paper* .

Klemperer, Paul. 1987. "Markets with Consumer Switching Costs." *The Quarterly Journal of Economics* 102 (2):375–394.

Lam, H. Tai. 2023. "Platform Search Design and Market Power." Working Paper.

Lee, Kwok-Hao and Leon Musolff. 2025. "Two-Sided Markets Shaped By Platform-Guided Search." Working Paper.

Leung, Tin Cheuk, Shi Qi, and Koleman S. Strumpf. 2026. "Dissecting Netflix's Self-Preferencing: Evidence from Viewer-Level Data." Working Paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6990898.

Leyden, Benjamin T. 2022. "There's an App (Update) for That: Understanding Product Updating Under Digitization." *Working Paper* .

———. 2025. "Platform design and innovation incentives: Evidence from the product rating system on Apple's App Store." *International Journal of Industrial Organization* 99:103133. URL https://www.sciencedirect.com/science/article/pii/S0167718724000882.

Li, Ding and Hsin-Tien Tsai. 2026. "Mobile Apps and Targeted Advertising: Competitive Effects of Data Sharing." *Journal of the European Economic Association* 24 (1):167–198. URL https://doi.org/10.1093/jeea/jvaf025.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb. 2023. "Cluster-Robust Inference: A Guide to Empirical Practice." *Journal of Econometrics* 232 (2):272–299.

Madsen, Erik and Nikhil Vellodi. 2025. "Insider Imitation." *Journal of Political Economy* 133 (2):652–709. URL https://doi.org/10.1086/732888.

Mills, David E. 1995. "Why Retailers Sell Private Labels." *Journal of Economics & Management Strategy* 4 (3):509–528.

Pichai, Sundar. 2017. "Keynote address, Google I/O 2017." Google I/O, Mountain View, CA. Two billion monthly active Android devices announced; reported by 9to5Google, https://9to5google.com/2017/05/17/numbers-google-io-2-billion-android/.

Raval, Devesh. 2023. "Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box." Working Paper.

Reimers, Nils and Iryna Gurevych. 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, 3982–3992. URL https://arxiv.org/abs/1908.10084.

Rosenthal, Robert W. 1980. "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price." *Econometrica* 48 (6):1575–1579.

Scott Morton, Fiona, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A Douglas Melamed, and Jamie Morgenstern. 2019. "Market Structure and Antitrust Subcommittee Report." In *Stigler Committee on Digital Platforms: Final Report*, edited by Luigi Zingales, Guy Rolnik, and Filippo Maria Lancieri. George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business, 23–138. URL https://research.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf.

Seamans, Robert and Feng Zhu. 2014. "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers." *Management Science* 60 (2):476–493.

Shaked, Avner and John Sutton. 1982. "Relaxing Price Competition Through Product Differentiation." *The Review of Economic Studies* 49 (1):3–13.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2):175–199. URL https://www.sciencedirect.com/science/article/pii/S030440762030378X. Themed Issue: Treatment Effect 1.

Teng, Xuan. 2026. "Self-preferencing, Quality Competition, and Welfare in Mobile Application Markets." Working Paper. URL https://sites.google.com/view/xuanteng/research.

Ward, Michael B., Jay P. Shimshack, Jeffrey M. Perloff, and J. Michael Harris. 2002. "Effects of the Private-Label Invasion in Food Industries." *American Journal of Agricultural Economics* 84 (4):961–973.

Warren, Elizabeth. 2019. "Here's how we can break up Big Tech." *Medium* URL https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c.

Wen, Wen and Feng Zhu. 2019. "Threat of platform-owner entry and complementor responses: Evidence from the mobile app market." *Strategic Management Journal* 40 (9):1336–1367.

Wooldridge, Jeffrey M. 2025. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." *Empirical Economics* 69 (5):2545–2587.

Zhu, Feng and Qihong Liu. 2018. "Competing with complementors: An empirical look at Amazon.com." *Strategic Management Journal* 39 (10):2618–2642.

Appendices

A Data Appendix

This appendix presents detailed summary statistics for the three analysis dimensions in the main text. Table 2 in the main text reports aggregate pre-treatment and post-treatment means; the tables below decompose these by market and by entry type.

A.1 Incumbent App Summary Statistics

Tables A.1 and A.2 present app-month level summary statistics for incumbent apps, broken down by market for the pre-treatment and post-treatment periods, respectively. These statistics correspond to Panel A of Table 2.

	Price	Price-Paid	Free	IAP	Avg. Rating	Update	Δ Log Ratings	N
Animoji	0.18	1.59	0.89	0.35	3.72	0.12	0.04	819
Apple Music 13.0	1.04	2.85	0.64	0.33	3.99	0.27	0.04	1035
Apple Research	1.67	6.05	0.72	0.24	4.09	0.26	0.05	261
Blood Oxygen	0.84	2.19	0.62	0.31	3.37	0.23	0.08	117
Breathe 10.0	0.00		1.00	0.29	4.19	0.14	0.04	1953
Camera 11.0	0.00		1.00	0.26	3.84	0.22	0.05	1755
Clock 10.0	0.00		1.00	0.25	4.17	0.18	0.04	2601
Development 11.0	0.00		1.00	0.29	3.85	0.20	0.03	2178
Ecg	2.73	3.87	0.29	0.25	3.88	0.20	0.04	459
Find Items	1.43	4.07	0.65	0.53	3.99	0.29	0.06	432
Health 10.0	0.00		1.00	0.21	3.87	0.19	0.03	252
Health 11.3	0.29	3.39	0.91	0.09	3.56	0.41	0.07	1656
Homekit 10.0	0.00		1.00	0.28	3.58	0.40	0.09	1287
Magnifier 10.0	0.00		1.00	0.25	3.74	0.13	0.04	585
Measure	3.15	3.92	0.20	0.24	3.92	0.18	0.05	711
Notes 11.0	0.00		1.00	0.30	4.00	0.24	0.04	3591
Radio 11.1	0.14	3.81	0.96	0.28	3.86	0.24	0.05	450
Reality Composer	2.76	5.02	0.45	0.46	3.91	0.19	0.05	441
Schoolwork	1.66	5.03	0.67	0.33	3.92	0.38	0.08	1368
Sleep	1.63	3.62	0.55	0.48	4.27	0.29	0.05	1548
Translate	1.79	4.65	0.62	0.59	4.26	0.24	0.04	4356
Walkie Talkie 12.0	1.08	3.52	0.69	0.46	3.79	0.33	0.05	1953
All Markets	0.75	4.09	0.82	0.35	3.98	0.25	0.05	29808

Table A.1: Incumbent App Summary Statistics by Market (Pre-Treatment)

A.2 Entry and Exit Summary Statistics

Tables A.3 and A.4 present market-platform-month level entry and exit statistics, broken down by market for the pre-treatment and post-treatment periods, respectively. These statistics use the market dynamics sample and correspond to Panel B of Table 2.

	Price	Price-Paid	Free	IAP	Avg. Rating	Update	Δ Log Ratings	N
Animoji	0.88	1.61	0.45	0.38	3.70	0.10	0.02	819
Apple Music 13.0	1.04	2.91	0.64	0.35	4.02	0.25	0.04	1035
Apple Research	1.65	5.99	0.72	0.24	4.13	0.31	0.02	261
Blood Oxygen	0.84	2.19	0.62	0.38	3.45	0.21	0.05	117
Breathe 10.0	0.00		1.00	0.30	4.18	0.12	0.02	1953
Camera 11.0	2.03	3.33	0.39	0.28	3.87	0.21	0.03	1755
Clock 10.0	0.00		1.00	0.28	4.16	0.16	0.02	2601
Development 11.0	3.70	5.38	0.31	0.31	3.85	0.22	0.03	2178
Ecg	2.86	4.05	0.29	0.27	3.85	0.09	0.03	459
Find Items	1.13	3.62	0.69	0.58	3.99	0.24	0.02	432
Health 10.0	0.00		1.00	0.26	3.88	0.19	0.01	252
Health 11.3	0.49	3.58	0.86	0.09	3.57	0.39	0.06	1656
Homekit 10.0	0.00		1.00	0.31	3.53	0.40	0.04	1287
Magnifier 10.0	0.00		1.00	0.27	3.72	0.12	0.02	585
Measure	3.12	3.94	0.21	0.27	3.94	0.16	0.04	711
Notes 11.0	2.76	4.64	0.41	0.32	4.02	0.24	0.03	3591
Radio 11.1	1.51	3.14	0.52	0.30	3.93	0.27	0.04	450
Reality Composer	2.69	4.89	0.45	0.47	3.82	0.17	0.02	441
Schoolwork	1.66	5.09	0.67	0.35	4.00	0.35	0.05	1368
Sleep	1.79	4.02	0.56	0.50	4.26	0.26	0.03	1548
Translate	1.79	4.72	0.62	0.60	4.25	0.20	0.02	4356
Walkie Talkie 12.0	1.08	3.49	0.69	0.47	3.80	0.33	0.03	1953
All Markets	1.53	4.24	0.64	0.37	3.98	0.23	0.03	29808

Table A.2: Incumbent App Summary Statistics by Market (Post-Treatment)

	Log Entry	Log Exit	N
Animoji	2.05	1.54	18
Apple Music 13.0	1.10	1.37	18
Apple Research	0.76	0.60	18
Blood Oxygen	0.45	0.29	18
Breathe 10.0	2.73	1.13	18
Camera 11.0	2.32	1.96	18
Clock 10.0	2.56	1.52	18
Development 11.0	1.76	1.51	18
Ecg	0.60	0.84	18
Find Items	0.53	0.51	18
Health 10.0	0.60	0.08	18
Health 11.3	1.82	1.33	18
Homekit 10.0	1.80	0.75	18
Magnifier 10.0	0.68	0.34	18
Measure	1.29	1.44	18
Notes 11.0	2.43	2.19	18
Radio 11.1	1.05	0.84	18
Reality Composer	0.98	1.00	18
Schoolwork	1.60	1.49	18
Sleep	1.54	1.51	18
Translate	2.16	2.55	18
Walkie Talkie 12.0	2.16	2.71	18
All Markets	1.50	1.25	396

Table A.3: Market-Level Entry and Exit by Market (Pre-Treatment)

	Log Entry	Log Exit	N
Animoji	1.79	1.86	18
Apple Music 13.0	0.71	1.15	18
Apple Research	0.65	0.52	18
Blood Oxygen	0.47	0.39	18
Breathe 10.0	2.58	1.64	18
Camera 11.0	2.21	2.21	18
Clock 10.0	2.63	1.98	18
Development 11.0	1.98	1.75	18
Ecg	0.59	0.62	18
Find Items	0.55	0.71	18
Health 10.0	0.57	0.50	18
Health 11.3	1.82	1.53	18
Homekit 10.0	2.01	0.96	18
Magnifier 10.0	0.70	0.88	18
Measure	1.03	1.03	18
Notes 11.0	2.15	2.28	18
Radio 11.1	0.96	0.86	18
Reality Composer	0.67	1.04	18
Schoolwork	1.50	1.61	18
Sleep	1.19	1.63	18
Translate	2.01	2.67	18
Walkie Talkie 12.0	1.89	2.39	18
All Markets	1.39	1.37	396

Table A.4: Market-Level Entry and Exit by Market (Post-Treatment)

A.3 Entering Cohort Summary Statistics

Tables A.5 and A.6 present summary statistics for the entering cohort analysis (repeated cross-section), broken down by market for the pre-treatment and post-treatment periods, respectively. These statistics use the market dynamics sample and correspond to Panel C of Table 2.

	Price	Price-Paid	Free	IAP	Avg. Rating	Log # Ratings	Similarity	N
Animoji	0.32	2.25	0.86	0.08	4.44	0.80	0.64	160
Apple Music 13.0	0.79	3.08	0.74	0.09	4.78	1.81	0.66	43
Apple Research	0.61	2.66	0.77	0.04	2.35	0.77	0.63	26
Blood Oxygen	1.30	4.24	0.69	0.15	4.73	1.73	0.66	13
Breathe 10.0	0.00		1.00	0.14	4.62	0.53	0.65	341
Camera 11.0	0.00		1.00	0.07	4.22	1.11	0.64	181
Clock 10.0	0.00		1.00	0.10	4.48	0.58	0.65	239
Development 11.0	0.00		1.00	0.05	3.52	0.79	0.63	114
Ecg	1.95	3.72	0.48	0.10	4.90	0.21	0.65	21
Find Items	1.05	2.21	0.53	0.37	4.64	1.45	0.65	19
Health 10.0	0.00		1.00	0.00	4.84	0.49	0.63	21
Health 11.3	0.09	2.19	0.96	0.03	4.39	0.39	0.65	116
Homekit 10.0	0.00		1.00	0.09	4.29	0.90	0.65	116
Magnifier 10.0	0.00		1.00	0.12	4.17	0.29	0.67	26
Measure	1.98	2.72	0.27	0.10	3.76	0.58	0.67	63
Notes 11.0	0.00		1.00	0.08	4.32	0.58	0.65	237
Radio 11.1	0.18	2.66	0.93	0.09	3.96	0.52	0.65	45
Reality Composer	2.42	4.99	0.51	0.23	4.36	1.53	0.64	35
Schoolwork	1.48	4.05	0.64	0.08	4.24	0.66	0.64	85
Sleep	0.99	2.69	0.63	0.32	4.60	1.12	0.66	76
Translate	0.74	3.57	0.79	0.55	4.32	0.95	0.68	178
Walkie Talkie 12.0	0.79	2.83	0.72	0.14	4.26	1.45	0.63	147
All Markets	0.37	3.15	0.88	0.14	4.33	0.80	0.65	2302

Table A.5: Entering Cohort Summary Statistics by Market (Pre-Treatment)

A.4 Summary Statistics by Entry Type

Table A.7 reports the statistics in Table 2 separately for standalone and integrated entry markets, defined in Section 2.2.

A.5 Google's Own Entry in the Apple-Entry Markets

A threat to the cross-platform control group would arise if Google entered the same submarkets as Apple at similar times, so that Play Store apps were themselves exposed to first-party entry during the event window. Table A.8 addresses this concern directly. For each of the 22 Apple-entry markets, it identifies the closest Google first-party product and records when that product became broadly available to the Android user base. Google entry dates, primary and archived sources, and per-market classification notes are recorded in a curated data file that accompanies the analysis code.

	Price	Price-Paid	Free	IAP	Avg. Rating	Log # Ratings	Similarity	N
Animoji	0.67	2.20	0.69	0.16	3.92	1.94	0.68	95
Apple Music 13.0	0.38	3.32	0.88	0.19	4.42	1.35	0.66	26
Apple Research	0.78	3.59	0.78	0.00	4.00	1.47	0.64	23
Blood Oxygen	1.12	3.59	0.69	0.12	4.25	1.61	0.70	16
Breathe 10.0	0.00		1.00	0.12	4.46	0.53	0.65	266
Camera 11.0	1.10	2.74	0.60	0.10	4.48	0.95	0.64	162
Clock 10.0	0.00		1.00	0.06	4.55	0.59	0.65	252
Development 11.0	1.51	3.45	0.56	0.10	4.23	1.25	0.64	153
Ecg	2.63	4.54	0.42	0.11	5.00	0.04	0.67	19
Find Items	0.78	3.49	0.78	0.44	4.44	1.48	0.66	18
Health 10.0	0.00		1.00	0.11	4.82	0.67	0.62	16
Health 11.3	0.35	3.32	0.89	0.04	4.82	0.49	0.64	113
Homekit 10.0	0.00		1.00	0.08	4.28	1.13	0.64	122
Magnifier 10.0	0.00		1.00	0.09	4.25	0.47	0.69	28
Measure	1.34	3.21	0.58	0.16	4.22	0.89	0.67	43
Notes 11.0	1.54	3.24	0.52	0.17	4.34	0.85	0.65	172
Radio 11.1	1.00	2.78	0.64	0.15	3.75	0.54	0.64	39
Reality Composer	1.54	4.62	0.67	0.29	4.37	0.91	0.65	24
Schoolwork	0.96	3.32	0.71	0.13	4.37	0.69	0.63	83
Sleep	0.95	2.91	0.67	0.51	4.45	1.38	0.65	49
Translate	0.78	2.85	0.73	0.44	4.33	1.24	0.67	131
Walkie Talkie 12.0	0.78	2.86	0.73	0.25	3.98	1.79	0.64	110
All Markets	0.68	3.11	0.78	0.16	4.31	0.95	0.65	1960

Table A.6: Entering Cohort Summary Statistics by Market (Post-Treatment)

The comparison is reassuring for the control group. In 9 of the 22 markets Google was a long-prior incumbent: its competing product—for example, Google Fit, Google Keep, Google Classroom, the Google Camera app, or Google Translate—predated Apple's entry by years, in several cases by close to a decade, so the Play Store panel reflects a stable competitive environment rather than a contemporaneous entry shock. In 8 markets Google was a follower, introducing a comparable product well after Apple and outside the event window. In 4 markets Google never shipped a first-party equivalent at all, as for the Apple Watch Walkie-Talkie feature and the face-tracked Animoji. Only 1 of the 22 markets shows a Google first-party product arriving within ±9 months of Apple's entry: ARCore, Google's closest Android analogue to Apple's ARKit, in the Development 11.0 market.

Even for this single market, contamination of the control group is unlikely to overturn the results. ARCore 1.0's launch was limited: it reached approximately 100 million Android smartphones across 13 supported device models in February 2018 (Gosalia, 2018), roughly five percent of the two billion monthly active Android devices Google reported in mid-2017 (Pichai, 2017), so the typical Play Store developer in the control group was not subject to a broad AR-platform-capability shock during the event window. More generally, if a concurrent Google entry affected Play Store developers in the same direction that Apple's entry affected App Store developers, the difference-in-differences estimator nets out part of that common effect and attenuates the estimated treatment effect toward

Table A.7: Summary Statistics by Entry Type

	Standalone			Integrated		
	Pre	Post	Change	Pre	Post	Change
Panel A: Incumbent Apps						
Price	1.768	1.835	+0.068	0.193	1.357	+1.164
Price (Paid Apps)	4.301	4.180	-0.120	3.286	4.281	+0.995
Free	0.589	0.561	-0.028	0.941	0.683	-0.258
In-App Purchases	0.462	0.484	+0.022	0.287	0.303	+0.016
Avg. Rating	4.100	4.102	+0.002	3.910	3.915	+0.005
Update	0.250	0.218	-0.032	0.246	0.244	-0.002
Δ Log # Ratings	0.048	0.026	-0.022	0.046	0.032	-0.015
N	10,512	10,512		19,296	19,296	
Panel B: Market-Level Entry and Exit						
Log Entry	1.196	1.044	-0.152	1.751	1.685	-0.067
Log Exit	1.175	1.208	+0.033	1.312	1.511	+0.199
N	180	180		216	216	
Panel C: Entering Cohorts						
Price	1.018	0.972	-0.046	0.104	0.582	+0.478
Price (Paid Apps)	3.247	3.122	-0.125	2.815	3.109	+0.294
Free	0.686	0.689	+0.002	0.963	0.813	-0.150
In-App Purchases	0.249	0.269	+0.021	0.093	0.119	+0.026
Avg. Rating	4.346	4.221	-0.125	4.325	4.351	+0.026
Log # Ratings	0.892	1.234	+0.342	0.757	0.859	+0.103
Similarity	0.655	0.663	+0.008	0.646	0.645	-0.001
N	676	501		1,626	1,459	

Notes: Panel A reports app-month level statistics for incumbent apps observed in an 18-month balanced panel around each entry event. Panel B reports market-platform-month level log entry and exit counts. Panel C reports statistics for apps observed once upon entry (repeated cross-section). Pre-Treatment refers to the 9 months before Apple's entry; Post-Treatment refers to the 9 months after.

zero. Concurrent Google entry therefore makes the headline estimates more conservative rather than inflating them.

Two of the classifications in Table A.8 warrant comment, as each involves a Google product that superficially appears concurrent but is not the relevant comparator. In the HomeKit market, Apple's Home app shipped with iOS 10 in September 2016, and Google renamed its existing Google Cast app to "Google Home" the following month. The renamed app carried over the Cast app's streaming-setup functionality but did not yet manage third-party smart-home accessories—the capability comparable to Apple's Home app, which Google added only in its October 2018 redesign—so I date Google's comparable entry to 2018 and classify it as a follower. In the Camera market, Google Lens launched in October 2017 but was a Pixel-2 exclusive until its broader Android rollout in March 2018. Under the requirement that a comparator be broadly available to the Android user base, the relevant Google product is instead the standalone Google Camera app, released in 2014, making Google a long-prior incumbent.

B Cohort Analysis Event Studies

This appendix presents event study estimates for the cohort composition analysis described in Section 3.1.2. I estimate:

$$y_j = \sum_{k \neq -1} \gamma_k \cdot \mathbf{1}[t_j - t_a^* = k] + \sum_{k \neq -1} \beta_k \cdot \mathbf{1}[t_j - t_a^* = k] \times \text{AS}_j + \delta_{ap} + \eta_t + \varepsilon_j, \tag{8}$$

where t_j is the month app j enters, t_a^* is Apple's entry month in market a, γ_k captures common relative-time effects across platforms, δ_{ap} are market×platform fixed effects, and η_t are calendar month fixed effects. AS_j indicates whether app j is on the App Store. Coefficients β_k show how the App Store–Play Store gap in entry characteristics evolves relative to Apple's entry timing.

Figure B.1 plots the estimated β_k from Equation (8) for the monetization outcomes and Figure B.2 for the quality and location outcomes.

C Market-Level Heterogeneity

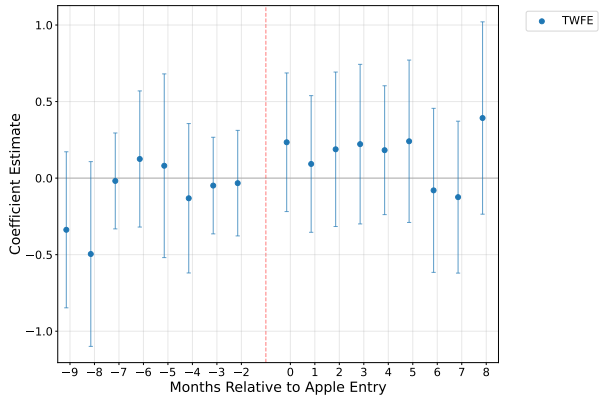
This appendix presents forest plots of market-specific treatment effects for all outcomes across the three margins of analysis. These figures complement the summary diagnostics reported in Table 6 and the selected forest plots presented in the main text (Figure 6). Figure C.1 covers the entry and exit outcomes, Figures C.2 and C.3 the entering-cohort outcomes, and Figures C.4 and C.5 the incumbent outcomes.

Table A.8: Google's First-Party Entry in the Apple-Entry Markets

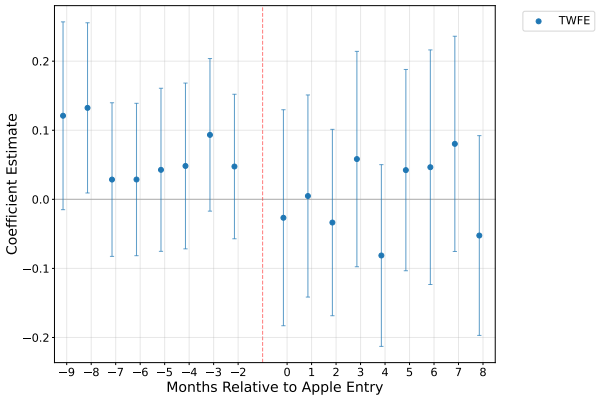
Market	Apple Entry	Google First-Party Equivalent	Google Entry	Gap (mo.)	In Window
Breathe 10.0	2016-09	Google Fit guided breathing (Wear OS)	2018-12	+27	No
Clock 10.0	2016-09	Google Clock Bedtime mode	2020-08	+47	No
Health 10.0	2016-09	Google Fit	2014-10	-23	No
Homekit 10.0	2016-09	Google Home app (October 2018 redesign)	2018-10	+25	No
Magnifier 10.0	2016-09	Android Magnification (accessibility)	2012-11	-46	No
Camera 11.0	2017-09	Google Camera app	2014-04	-41	No
Development 11.0	2017-09	ARCore 1.0	2018-02	+5	Yes
Notes 11.0	2017-09	Google Keep	2013-03	-54	No
Radio 11.1	2017-10	Google Play Music (radio/stations)	2011-11	-71	No
Animoji	2017-11	No first-party equivalent	–	–	No
Health 11.3	2018-03	No first-party equivalent	–	–	No
Schoolwork	2018-06	Google Classroom	2014-08	-46	No
Measure	2018-09	Google Measure (Project Tango)	2016-11	-22	No
Walkie Talkie 12.0	2018-09	No first-party equivalent	–	–	No
Ecg	2018-12	Pixel Watch ECG app	2022-10	+46	No
Apple Music 13.0	2019-09	Google Play Music / YouTube Music	2011-11	-94	No
Reality Composer	2019-09	No first-party equivalent	–	–	No
Apple Research	2019-11	Google Health Studies	2020-12	+13	No
Sleep	2020-09	Pixel Watch sleep tracking (Fitbit)	2022-10	+25	No
Translate	2020-09	Google Translate app	2010-01	-128	No
Blood Oxygen	2020-10	Pixel Watch (blood oxygen / SpO2)	2023-06	+32	No
Find Items	2021-04	Google Find My Device network (item trackers)	2024-04	+36	No

Notes: Each row pairs an Apple-entry market with the closest Google first-party product. Apple entry dates are the registry release dates used throughout the paper. Google entry dates are the dates each Google product became broadly available to the Android user base; full dates, primary and archived sources, and per-market notes are recorded in the curated Google-entries data file. Gap is the number of months between Google's entry and Apple's entry, with negative values indicating that Google entered first. A market is flagged as in-window when the gap is within ±9 months, the half-length of the paper's balanced-panel event window (estimation window $[-9, +8]$). Under a strict attribution rule, a product counts as Google only if Google-branded: Fitbit features count as Google from the close of the acquisition (January 2021), and Alphabet sister-companies are excluded.

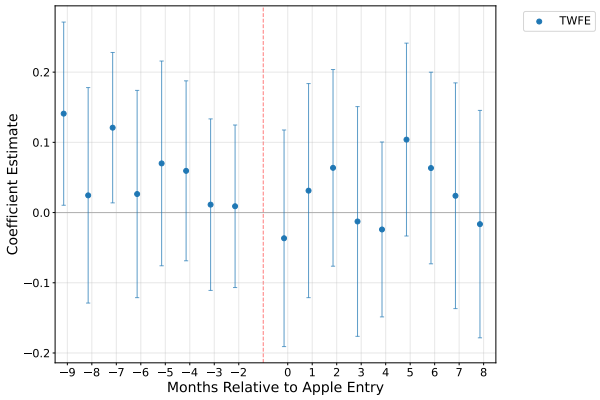
Figure B.1: Cohort Event Studies: Monetization Outcomes



(a) Price (All Apps)



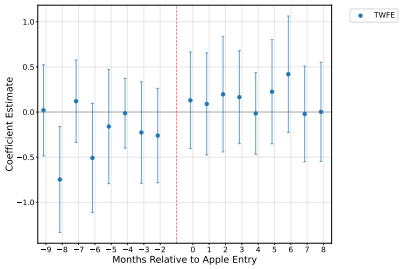
(b) Free



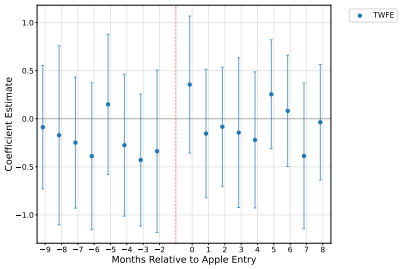
(c) In-App Purchases

Note: Event study estimates from the cross-sectional DiD model showing how the App Store–Play Store gap in monetization characteristics evolves relative to Apple's entry timing. Each app is observed once at entry. Standard errors are clustered at the market-platform level. Error bars represent 95% confidence intervals.

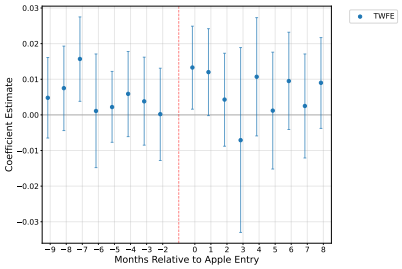
Figure B.2: Cohort Event Studies: Quality and Location Outcomes



(a) Rating



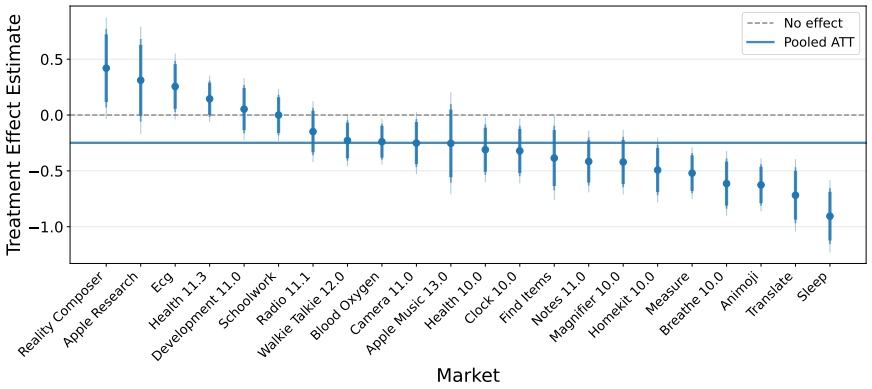
(b) Log Rating Count



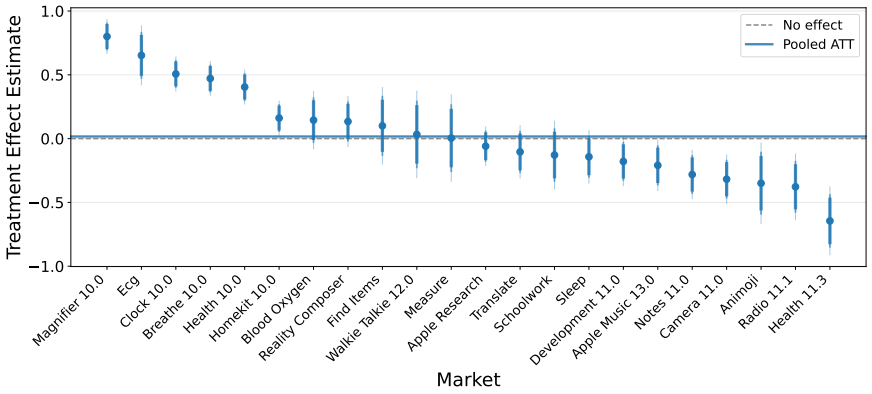
(c) Similarity

Note: Event study estimates from the cross-sectional DiD model showing how the App Store–Play Store gap in quality and location characteristics evolves relative to Apple's entry timing. Each app is observed once at entry. Standard errors are clustered at the market-platform level. Error bars represent 95% confidence intervals.

Figure C.1: Market Heterogeneity in Entry and Exit



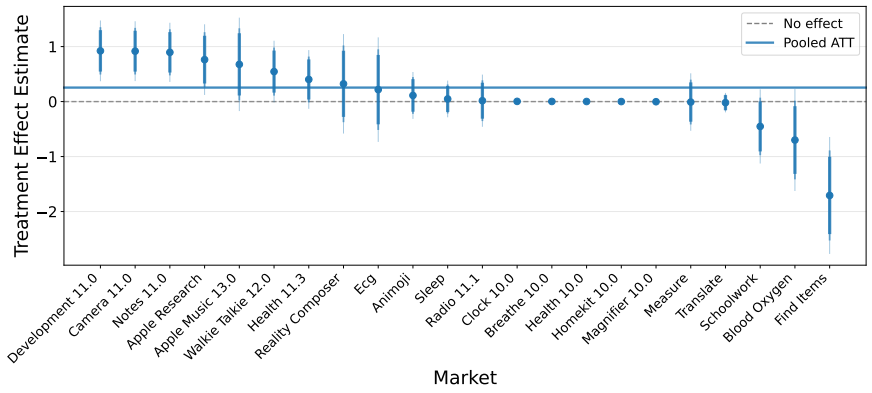
(a) Log Entry Count



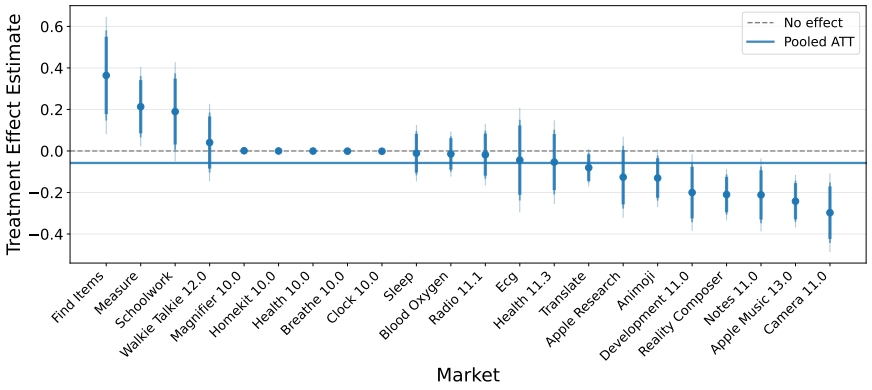
(b) Log Exit Count

Note: Market-specific treatment effects from the entry/exit model, Equation (2). Analysis uses the market dynamics sample at the market-platform-month level. Standard errors are clustered at the market-platform level. Points show estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines).

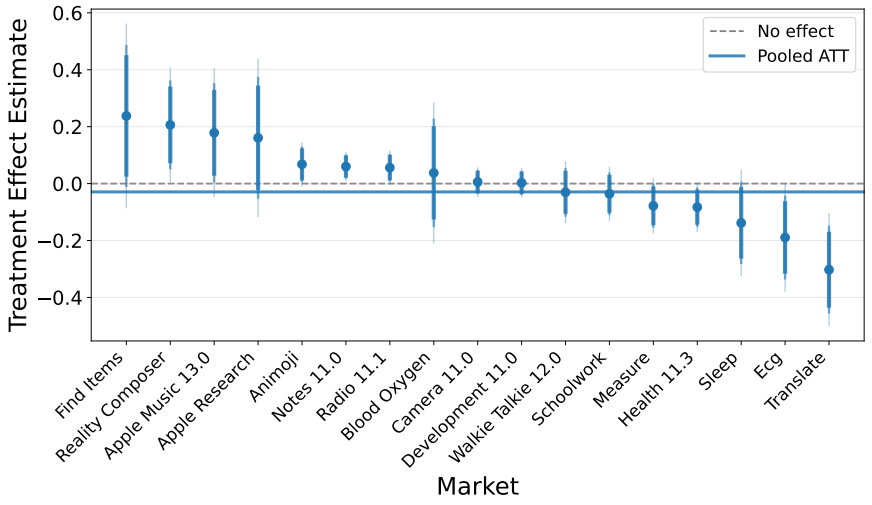
Figure C.2: Market Heterogeneity in Cohort Monetization



(a) Price



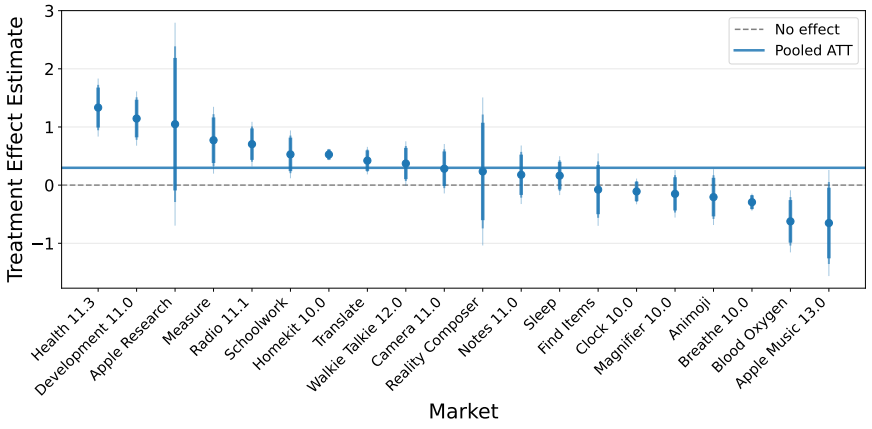
(b) Free



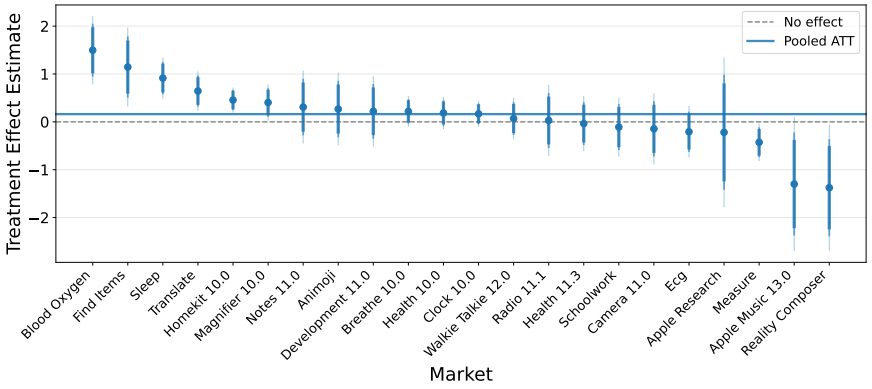
(c) In-App Purchases

Note: Market-specific treatment effects from the cohort composition model, Equation (3). Analysis uses the market dynamics sample. Each app is observed once at entry. Standard errors are clustered at the market-platform level. Points show estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines).

Figure C.3: Market Heterogeneity in Cohort Quality



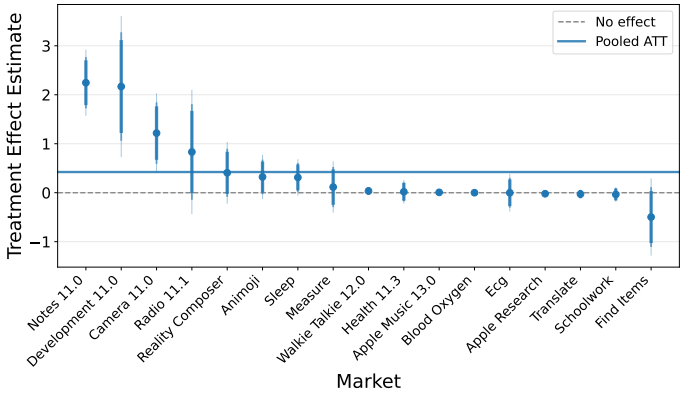
(a) Avg. Rating



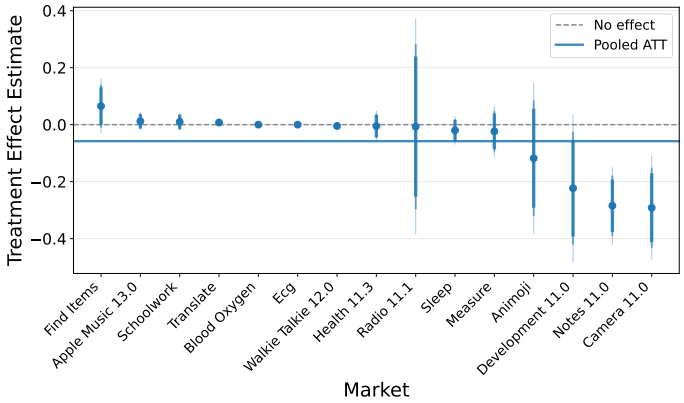
(b) Log Rating Count

Note: Market-specific treatment effects from the cohort composition model, Equation (3). Analysis uses the market dynamics sample. Each app is observed once at entry. Standard errors are clustered at the market-platform level. Points show estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines).

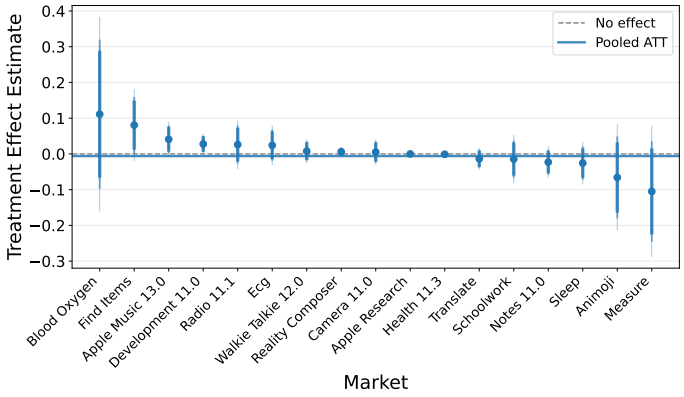
Figure C.4: Incumbent Market Heterogeneity in Monetization



(a) Price (All Apps)



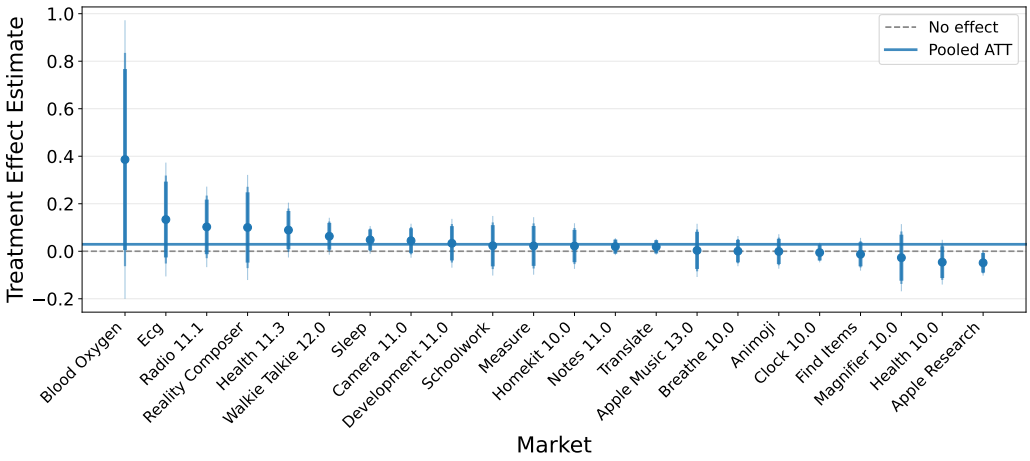
(b) Free



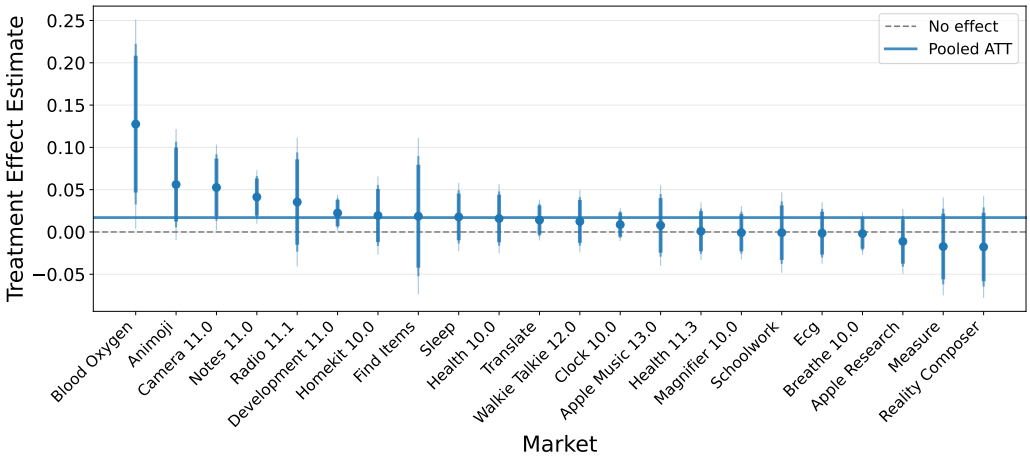
(c) In-App Purchases

Note: Market-specific treatment effects from the incumbent analysis model, Equation (4), at the app-month level. Markets lacking sufficient within-market variation are omitted; the IAP outcome additionally excludes 5 markets with sparse pre-period Android in-app purchase data. Standard errors are clustered at the app level. Points show estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines).

Figure C.5: Incumbent Market Heterogeneity in Quality



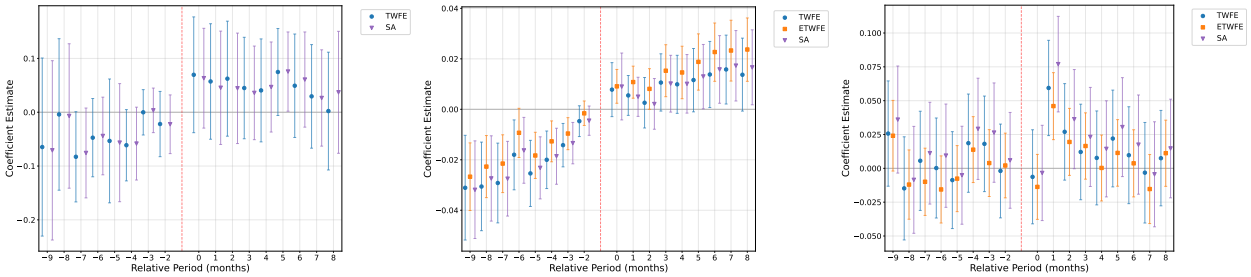
(a) Rating



(b) Rating Count Growth

Note: Market-specific treatment effects from the incumbent analysis model, Equation (4). Analysis is at the app-month level. Markets with insufficient outcome variation are omitted. Standard errors are clustered at the app level. Points show estimates with 90%, 95%, and 99% confidence intervals (thick to thin lines).

Figure D.1: Incumbent Treatment Effect Dynamics: Additional Outcomes



(a) Price (Paid Apps Only) (b) Rating (c) Update Frequency

Note: Event study estimates from the incumbent analysis model, Equation (4). TWFE and SA include app and market×month fixed effects. ETWFE (Wooldridge, 2025) uses app and month fixed effects; its saturated market×time treatment interactions already absorb market-specific time variation. SA (Sun and Abraham, 2021) shown for comparison. Analysis is at the app-month level. ETWFE is excluded from Price (Paid Apps) because insufficient within-market price variation prevents reliable estimation of market-specific treatment effects. Standard errors are clustered at the app level. Error bars represent 95% confidence intervals.

D Robustness

This appendix presents supporting material for the main analysis: event studies for the incumbent outcomes not shown in Figure 5, a comparison across DiD estimators for both incumbent and entry/exit results, an alternative fixed effects specification, alternative clustering of standard errors, robustness to the market-definition threshold, robustness to anticipation of Apple's entry, and robustness to alternative sample restrictions.

D.1 Additional Incumbent Outcome Dynamics

Figure D.1 reports event studies for the incumbent outcomes omitted from Figure 5 in the main text—paid-app price, average rating, and update frequency—estimated under the primary app and market×month fixed effects specification.

D.2 Comparison Across DiD Estimators

This subsection compares results across multiple difference-in-differences estimators to assess robustness. Alongside the primary TWFE estimator, I report results from the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) and the interaction-weighted estimator of Sun and Abraham (2021), both of which provide consistent estimates under heterogeneous treatment effects by avoiding the use of already-treated units as controls.

Incumbent Analysis Robustness. Table D.1 presents estimates across all three estimators for incumbent app outcomes. Results are qualitatively similar across methods: all three estimators agree on the sign of every effect, and on significance for the headline price, free/paid, and rating

Table D.1: Estimator Comparison: Incumbent App Outcomes

Outcome	(1) TWFE	(2) ETWFE	(3) SA
Δ Log Rating Count	0.0170***	0.0067***	0.0033
	(0.0035)	(0.0025)	(0.0032)
Free App	-0.0580***	-0.1007***	-0.0725***
	(0.0092)	(0.0068)	(0.0112)
In-App Purchases	-0.0061	-0.0032	-0.0062
	(0.0055)	(0.0026)	(0.0043)
Price	0.4221***	0.6969***	0.5465***
	(0.0464)	(0.0413)	(0.0549)
Price (Paid Apps)	0.0842*	—	0.0485
	(0.0494)	—	(0.0400)
Avg. Rating	0.0293***	0.0163***	0.0110**
	(0.0069)	(0.0043)	(0.0051)
Update	0.0104	0.0088	0.0229
	(0.0077)	(0.0096)	(0.0143)

Notes: Analysis is at the app-month level. Columns show treatment effect estimates from TWFE (two-way fixed effects), ETWFE (extended two-way fixed effects; Wooldridge (2025)), and SA (interaction-weighted estimator; Sun and Abraham (2021)). For the IAP outcome, 5 markets are excluded due to insufficient Android in-app purchase data in the pre-period. Standard errors clustered at the app level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

outcomes. ETWFE and Sun–Abraham tend to produce somewhat larger point estimates for the monetization outcomes; the rating-count-growth effect is positive under all three but not statistically significant under Sun–Abraham. Because the incumbent specification absorbs market×month effects and identifies the treatment from within-market comparisons against a never-treated Play Store control, differences across estimators here reflect their weighting of heterogeneous market effects rather than the negative-weighting bias of staggered designs (see Section 3.1.3). TWFE and SA use the primary app and market×month fixed effects specification. ETWFE uses app and month fixed effects. In this implementation, each market constitutes its own cohort, so the saturated market×time treatment interactions absorb market-specific time variation, making separate market×month fixed effects redundant. The ETWFE estimate for Price (Paid Apps) is excluded due to insufficient within-market variation in paid app prices, which prevents reliable estimation of market-specific treatment effects.

Entry/Exit Analysis Robustness. Table D.2 presents market-level entry and exit results across estimators using the same market dynamics sample as the main text (Table 3). Entry

Table D.2: Estimator Comparison: Market Entry and Exit

Outcome	(1) TWFE	(2) ETWFE	(3) SA
Log Entry Count	-0.2487***	-0.1553	-0.1553
	(0.0756)	(0.4124)	(0.1198)
Log Exit Count	0.0167	0.0967	0.0967
	(0.0718)	(0.3971)	(0.1108)

Notes: Analysis is at the market-platform-month level. Columns show treatment effect estimates from TWFE (two-way fixed effects), ETWFE (extended two-way fixed effects; Wooldridge (2025)), and SA (interaction-weighted estimator; Sun and Abraham (2021)). Standard errors clustered at the market-platform level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

effects are negative across all estimators, consistent with the entry deterrence documented in the main text, though the heterogeneity-robust estimators yield a smaller and less precisely estimated entry effect that is not significant at conventional levels. Exit effects are small and imprecisely estimated, consistent with the null average effect reported in the main text. All three estimators agree on the sign of both effects. Because this market-level specification uses calendar-month rather than market×month fixed effects (Section 3.1.1), I do not claim that the raw TWFE estimate is free of the negative weighting that can affect staggered designs; the heterogeneity-robust estimators instead provide the safeguard, avoiding the use of already-treated markets as controls by construction. That they retain the negative entry sign—attenuating its magnitude and precision rather than reversing it—indicates the entry-deterrence result is not an artifact of negative weighting, even as it tempers how precisely the pooled effect is estimated.

D.3 Alternative Fixed Effects Specification

The main text uses app and market×month fixed effects ($\gamma_j + \eta_{a(j),t}$) to absorb market-specific time shocks. This subsection presents results using the simpler app and month fixed effects ($\gamma_j + \eta_t$) as a robustness check. Results are directionally consistent with the main specification. The unconditional price and the share of free apps are larger in magnitude under app and month fixed effects, while the other five outcomes are somewhat smaller, and all are more precisely estimated.

Monetization Outcomes

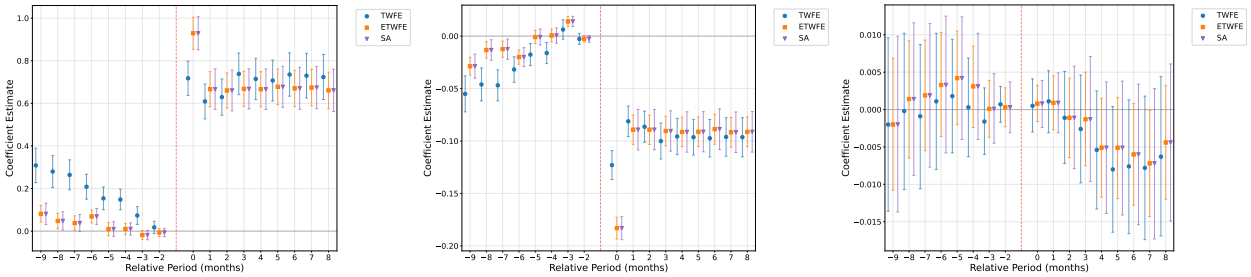
Table D.3 reports the pooled ATTs and Figure D.2 the event-study dynamics for the monetization outcomes under app and month fixed effects.

Table D.3: Incumbent App Monetization Outcomes (App + Month FE)

	(1) Price	(2) Price (Paid Apps)	(3) Free App	(4) In-App Purchases
ATT	0.5623***	0.0553*	-0.0800***	-0.0030
	(0.0432)	(0.0328)	(0.0079)	(0.0042)
Baseline Mean	0.7484	4.090	0.8170	0.3704
N	59,616	16,187	59,616	46,260

Notes: Estimates from the incumbent analysis model at the app-month level. For the IAP outcome, 5 markets are excluded due to insufficient Android in-app purchase data in the pre-period. Standard errors clustered at the app level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Figure D.2: Monetization Dynamics (App + Month FE)



(a) Price (All Apps) (b) Free (c) In-App Purchases

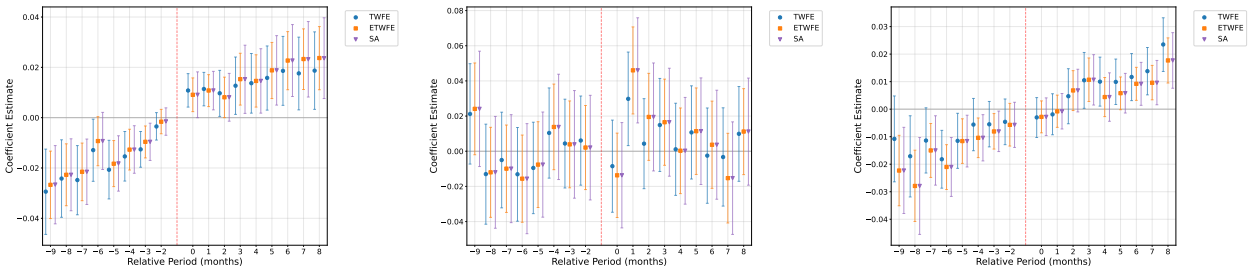
Note: Event study estimates using app and month fixed effects. TWFE, ETWFE (Wooldridge, 2025), and SA (Sun and Abraham, 2021) are shown. Analysis is at the app-month level. For the IAP outcome, 5 markets are excluded due to insufficient Play Store in-app purchase data in the pre-period. Standard errors are clustered at the app level. Error bars represent 95% confidence intervals.

Table D.4: Incumbent App Quality Outcomes (App + Month FE)

	(1) Avg. Rating	(2) Update	(3) Δ Log Rating Count
ATT	0.0269***	0.0065	0.0148***
	(0.0059)	(0.0069)	(0.0032)
Baseline Mean	3.978	0.2475	0.0470
N	56,770	59,616	59,616

Notes: Estimates from the incumbent analysis model at the app-month level. Standard errors clustered at the app level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Figure D.3: Quality Dynamics (App + Month FE)



(a) Rating (b) Update Frequency (c) Rating Count Growth

Note: Event study estimates using app and month fixed effects. TWFE, ETWFE (Wooldridge, 2025), and SA (Sun and Abraham, 2021) are shown. Analysis is at the app-month level. Standard errors are clustered at the app level. Error bars represent 95% confidence intervals.

Quality Outcomes

Table D.4 and Figure D.3 report the corresponding results for the quality outcomes.

D.4 Clustering Robustness

All incumbent app-level regressions cluster standard errors at the app level. This choice addresses within-app serial correlation (Bertrand, Duflo, and Mullainathan, 2004) and provides reliable asymptotic inference given the large number of clusters (thousands of apps). Because the analysis covers the full population of apps on both platforms—rather than a sample drawn from a larger population—the design-based motivation for coarser clustering is weaker than in sampled settings (Abadie et al., 2023).

Nevertheless, since treatment is assigned at the market level (with 22 affected markets, yielding approximately 44 market-platform cells), one may be concerned about within-market residual correlation among apps. Standard market-level cluster-robust variance estimators (CRV1) are known to be downward biased with so few clusters (Hansen, 2025; MacKinnon, Nielsen, and Webb, 2023). Instead, I report CRV3J (cluster jackknife) standard errors clustered at the market-platform level,

Table D.5: Clustering Robustness: App-Level vs. Market-Platform CRV3J Standard Errors

	ATT	App-Level SE	CRV3J SE
Price	0.422	(0.046)***	(0.368)
Price (Paid Apps)	0.084	(0.049)*	(0.102)
Free	-0.058	(0.0092)***	(0.051)
In-App Purchases	-0.0061	(0.0055)	(0.0093)
Rating	0.029	(0.0069)***	(0.011)***
Update	0.010	(0.0077)	(0.013)
Δ Log Rating Count	0.017	(0.0035)***	(0.0065)***

Notes: Column "ATT" reports the average treatment effect on the treated from the TWFE specification (Equation (4)). Column "App-Level SE" reports standard errors clustered at the app level (primary specification). Column "CRV3J SE" reports cluster-jackknife standard errors (CRV3J) clustered at the market-platform level, which guard against downward bias with few clusters (MacKinnon, Nielsen, and Webb, 2023; Hansen, 2025). Significance stars on each SE column use the respective inference. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

which are never downward biased by construction.

Table D.5 presents the comparison. For each outcome, I report the ATT estimate alongside both the primary app-level standard errors and the CRV3J standard errors. The CRV3J standard errors are larger for most outcomes, as expected given that only 44 market-platform clusters underlie the variance estimate. For quality outcomes (rating and rating count growth), which display broadly distributed effects across markets, significance is maintained under CRV3J inference. For monetization outcomes such as price and free, the CRV3J standard errors are substantially larger and these average effects are no longer statistically significant at conventional levels. This reflects both the small number of market-platform clusters and the concentration of pricing effects in a subset of markets—consistent with the market heterogeneity analysis in Section 3.2.

D.5 Market-Definition Threshold Robustness

The main analysis uses a cosine-similarity threshold of $\theta = 0.6$ to define markets, chosen to yield tight, focused competitive sets (Section 2.4). Because this single parameter determines which apps are treated as competitors, a natural concern is that the paper's conclusions may depend on this choice. This subsection re-estimates the primary TWFE pooled ATTs across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$ for all three margins—incumbent apps, entering cohorts, and market-level entry/exit—and re-runs the standalone/integrated comparison at each threshold.

Market size is monotone decreasing in θ, and at the tightest threshold some markets have too few apps meeting the similarity cutoff to remain in the estimation sample: at $\theta = 0.7$, 2 markets drop from the incumbent-app panel (health_10-0, apple_research), and none drop at any looser

Table D.6: Threshold Robustness (Incumbent Apps): TWFE Pooled ATTs Across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$

Outcome	(1) $\theta = 0.50$	(2) $\theta = 0.55$	(3) $\theta = 0.60$	(4) $\theta = 0.65$	(5) $\theta = 0.70$
Price	0.2758***	0.2745***	0.4221***	0.3284***	0.2613**
	(0.0360)	(0.0382)	(0.0464)	(0.0685)	(0.1276)
In-App Purchases	-0.0061*	-0.0070*	-0.0061	-0.0113	-0.0274
	(0.0035)	(0.0039)	(0.0055)	(0.0091)	(0.0188)
Avg. Rating	0.0346***	0.0339***	0.0293***	0.0269***	0.0266
	(0.0049)	(0.0054)	(0.0069)	(0.0101)	(0.0181)
Update	0.0120**	0.0123**	0.0104	0.0092	-0.0002
	(0.0049)	(0.0056)	(0.0077)	(0.0127)	(0.0234)
Δ Log Rating Count	0.0146***	0.0143***	0.0170***	0.0127*	0.0222
	(0.0021)	(0.0024)	(0.0035)	(0.0067)	(0.0142)
# Apps	7,060	5,650	3,312	1,455	522
# Markets	22	22	22	22	20
# Dropped	0	0	0	0	2

Notes: TWFE pooled ATTs for incumbent app outcomes. All columns use app and market×month fixed effects. Standard errors (clustered at the app level) in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
The set of active markets and severity-based exclusions are held fixed at their $\theta = 0.6$ values; pretrend diagnostics are not re-derived per threshold.

threshold. The exhibits below report ATTs computed on the effective (non-empty) market set at each θ; table footers show the effective market count per column.

Incumbent app outcomes. Table D.6 reports pooled ATTs for the five incumbent outcomes at each threshold. Four of the five keep the same sign at every θ; the update effect is small throughout, significant at the 5% level only at the two loosest thresholds ($\theta = 0.5$ and $\theta = 0.55$), indistinguishable from zero at the $\theta = 0.6$ baseline and tighter, and slightly negative at $\theta = 0.7$. Price and Avg. Rating are significant at the 1% level at every threshold from $\theta = 0.5$ through $\theta = 0.65$, and Δ Log Rating Count at the 1% level through $\theta = 0.6$ and the 10% level at $\theta = 0.65$; Price remains significant at $\theta = 0.7$, while Avg. Rating and Δ Log Rating Count lose significance there. Point estimates are stable across the range: the Price effect moves between \$0.26 and \$0.42, and the rating-count growth effect between 0.013 and 0.022 log points. At $\theta = 0.7$ the incumbent sample shrinks from 3,312 apps at the $\theta = 0.6$ baseline to 522 and standard errors widen by a factor of two to four, which explains the loss of precision on Avg. Rating and Δ Log Rating Count at that threshold rather than any material change in the underlying point estimates—if anything, the Δ Log Rating Count estimate at $\theta = 0.7$ (0.022) is slightly larger than at the $\theta = 0.6$ baseline (0.017), simply estimated far less precisely.

Table D.7: Threshold Robustness (Entering Cohorts): TWFE Pooled ATTs Across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$

Outcome	(1) $\theta = 0.50$	(2) $\theta = 0.55$	(3) $\theta = 0.60$	(4) $\theta = 0.65$	(5) $\theta = 0.70$
Price	0.1196	0.1203	0.2544**	0.1390	-0.3230
	(0.1089)	(0.1067)	(0.1261)	(0.1822)	(0.2152)
In-App Purchases	0.0042	-0.0096	-0.0292	0.0260	0.0036
	(0.0242)	(0.0332)	(0.0373)	(0.0404)	(0.0861)
Avg. Rating	0.1691	0.1905*	0.2978**	0.4123**	0.0886
	(0.1125)	(0.1113)	(0.1409)	(0.1743)	(0.2197)
Log Rating Count	0.1705**	0.1626**	0.1613	0.2358	0.3328
	(0.0798)	(0.0807)	(0.1326)	(0.2330)	(0.4334)
Similarity	-0.0010	0.0006	0.0016	0.0060	0.0002
	(0.0024)	(0.0029)	(0.0051)	(0.0061)	(0.0050)
# Entering Apps	12,315	8,568	4,262	1,593	502
# Markets	22	22	22	22	22
# Dropped	0	0	0	0	0

Notes: TWFE pooled ATTs for entering-cohort outcomes. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
The set of active markets and severity-based exclusions are held fixed at their $\theta = 0.6$ values; pretrend diagnostics are not re-derived per threshold.

Entering-cohort outcomes. Table D.7 reports the cohort results, which are less stable across the sweep than the incumbent estimates: no cohort outcome is significant at the 5% level at all five thresholds. Of the two effects the main text highlights, the rating improvement is the more robust. Avg. Rating is significant at the 5% level at the $\theta = 0.6$ baseline (0.30 stars) and at $\theta = 0.65$ (0.41 stars), and at the 10% level at $\theta = 0.55$, while the entering-cohort price increase reaches significance only at the baseline and is not reproduced at adjacent thresholds. Log Rating Count is significant only at the two loosest thresholds, $\theta = 0.5$ and $\theta = 0.55$, where the entering sample is largest—consistent with the main text's reading of the pooled rating-count effect as positive but insignificant. In-app purchases and similarity are statistically indistinguishable from zero at every threshold. The sweep leaves the qualitative cohort conclusions of the main text unchanged.

Market entry and exit. Table D.8 reports entry and exit results. The entry effect is negative and statistically significant from $\theta = 0.5$ through $\theta = 0.65$, with point estimates between -0.20 and -0.25. At $\theta = 0.7$ it attenuates to -0.04 and is no longer significant. The exit effect is small and statistically insignificant at every threshold except $\theta = 0.5$, where it is negative and significant. Its sign is not stable across thresholds, consistent with the null average exit effect in the main text. At $\theta = 0.7$, 3 markets (health_10-0, development_11-0, animoji) drop out of the entry/exit panel because the similarity cutoff is strict enough that too few apps remain to define those markets; the

Table D.8: Threshold Robustness (Market Entry & Exit): TWFE Pooled ATTs Across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$

Outcome	(1) $\theta = 0.50$	(2) $\theta = 0.55$	(3) $\theta = 0.60$	(4) $\theta = 0.65$	(5) $\theta = 0.70$
Log Entry Count	-0.2074***	-0.2104***	-0.2487***	-0.2014**	-0.0357
	(0.0580)	(0.0598)	(0.0756)	(0.0792)	(0.0601)
Log Exit Count	-0.1767**	-0.0622	0.0167	-0.0430	0.0498
	(0.0794)	(0.0819)	(0.0718)	(0.0711)	(0.0536)
# Markets	22	22	22	22	19
# Months	73	73	73	73	73
# Dropped	0	0	0	0	3

Notes: TWFE pooled ATTs for market-level entry/exit outcomes. Analysis is at the market-platform-month level. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
The set of active markets and severity-based exclusions are held fixed at their $\theta = 0.6$ values; pretrend diagnostics are not re-derived per threshold.

attenuation and loss of precision at this threshold reflect a market definition drawn too narrowly to be informative, not a reversal of the entry-deterrence result, which holds at every threshold through $\theta = 0.65$.

Standalone versus integrated entry effects across thresholds. The final exercise re-runs the standalone/integrated comparison from Table 7 at each threshold. Table D.9 reports the integrated and standalone pooled ATTs and their difference for all seven incumbent outcomes, with the $\theta = 0.6$ column reproducing the main-text estimates. At every threshold from $\theta = 0.5$ through $\theta = 0.65$, integrated entries raise prices by \$0.33 to \$0.54 more than standalone entries and lower the share of free apps by 5.0 to 10.7 percentage points more, with both differences significant at $p < 0.01$. At $\theta = 0.7$ both differentials retain their signs—\$0.48 for Price and 8.6 percentage points for Free—but are significant only at the 10% level, reflecting the sample contraction documented above: just 522 incumbent apps meet the similarity cutoff at that threshold.

D.6 Anticipation Robustness

Apple's release cycle creates a systematic gap between when developers learn about an entry and when it ships: major operating-system features are often announced at the Worldwide Developers Conference (WWDC) in June and released with the iOS update in September, three months later. Of the 22 markets in the sample, 18 were publicly previewed by Apple before release, with the announcement venue varying across Apple's annual WWDC keynotes, September iPhone keynotes, spring product events, and an Apple Newsroom press release. The median announcement-to-release lead is three months. If developers respond to announcements rather than waiting for the actual release, then the months immediately before entry will reflect treatment rather than counterfactual

Table D.9: Threshold Robustness (Entry Type): Standalone vs. Integrated TWFE ATTs Across $\theta \in \{0.5, 0.55, 0.6, 0.65, 0.7\}$

Outcome	(1) $\theta = 0.50$	(2) $\theta = 0.55$	(3) $\theta = 0.60$	(4) $\theta = 0.65$	(5) $\theta = 0.70$
Panel A: Integrated Entry					
Price	0.391*** (0.0534)	0.362*** (0.0552)	0.586*** (0.0677)	0.551*** (0.1137)	0.505** (0.2447)
Free	-0.054*** (0.0085)	-0.048*** (0.0093)	-0.076*** (0.0135)	-0.106*** (0.0224)	-0.074* (0.0446)
In-App Purchases	-0.0021 (0.0034)	-0.0034 (0.0042)	-0.0008 (0.0042)	0.0033 (0.0049)	0.016 (0.0141)
Price (Paid Apps)	0.050 (0.0511)	0.0052 (0.0327)	0.016 (0.0494)	0.119** (0.0537)	0.167 (0.1047)
Update	0.016 (0.0109)	0.024* (0.0124)	0.024 (0.0177)	-0.0075 (0.0308)	0.022 (0.0578)
Rating	0.016*** (0.0045)	0.015*** (0.0050)	0.0094 (0.0062)	0.0027 (0.0090)	0.021 (0.0158)
Δ Log Rating Count	0.0029 (0.0031)	0.0046 (0.0034)	0.0038 (0.0045)	0.0022 (0.0084)	0.0099 (0.0176)
Panel B: Standalone Entry					
Price	0.027 (0.0163)	0.029 (0.0187)	0.049* (0.0286)	0.052 (0.0383)	0.022 (0.0365)
Free	0.0009 (0.0023)	0.0015 (0.0024)	-0.0013 (0.0040)	0.0018 (0.0056)	0.013* (0.0068)
In-App Purchases	-0.0027 (0.0036)	-0.0066 (0.0044)	-0.011 (0.0071)	-0.016 (0.0116)	-0.039 (0.0255)
Price (Paid Apps)	0.055 (0.0349)	0.060 (0.0379)	0.056 (0.0551)	0.081 (0.0737)	0.178 (0.1324)
Update	0.0050 (0.0138)	0.0019 (0.0164)	-0.0002 (0.0222)	-0.0054 (0.0337)	-0.135** (0.0552)
Rating	0.017*** (0.0058)	0.017** (0.0066)	0.011 (0.0084)	0.0071 (0.0125)	-0.022 (0.0222)
Δ Log Rating Count	-0.0030 (0.0039)	-0.0013 (0.0042)	-0.0013 (0.0041)	-0.0057 (0.0068)	0.0050 (0.0112)
Panel C: Difference (Integrated − Standalone)					
Price	0.364*** (0.0558)	0.333*** (0.0583)	0.537*** (0.0735)	0.499*** (0.1200)	0.483* (0.2474)
Free	-0.055*** (0.0088)	-0.050*** (0.0096)	-0.075*** (0.0141)	-0.107*** (0.0231)	-0.086* (0.0451)
In-App Purchases	0.0005 (0.0050)	0.0032 (0.0061)	0.010 (0.0083)	0.020 (0.0126)	0.054* (0.0291)
Price (Paid Apps)	-0.0050 (0.0619)	-0.055 (0.0500)	-0.040 (0.0740)	0.037 (0.0912)	-0.011 (0.1688)
Update	0.011 (0.0176)	0.022 (0.0205)	0.024 (0.0283)	-0.0021 (0.0456)	0.156* (0.0800)
Rating	-0.0009 (0.0074)	-0.0014 (0.0083)	-0.0020 (0.0105)	-0.0044 (0.0154)	0.043 (0.0273)
Δ Log Rating Count	0.0059 (0.0049)	0.0059 (0.0054)	0.0051 (0.0061)	0.0079 (0.0108)	0.0050 (0.0208)
# Apps	7,060	5,650	3,312	1,455	522
# Markets	22	22	22	22	20
# Dropped	0	0	0	0	2

Notes: TWFE entry-type interaction estimates for incumbent app outcomes. Panels A and B report the pooled ATT for markets Apple entered via OS integration and via a standalone app, respectively; Panel C reports their difference (Integrated − Standalone). All columns use app and market×month fixed effects. Standard errors (clustered at the app level) in parentheses. *** p<0.01, ** p<0.05, * p<0.10. The $\theta = 0.60$ column corresponds to the main-text specification.
The set of active markets and severity-based exclusions are held fixed at their $\theta = 0.6$ values; pretrend diagnostics are not re-derived per threshold.

behavior, contaminating the pre-trend window and potentially biasing the pooled ATT estimates reported in the main text.

I assess robustness to this concern with two independent specifications. The *announcement-shifted* specification reassigns each market's treatment date to the earliest public Apple announcement date and re-estimates the TWFE pooled ATT under this earlier anchor. I recompute event time and the never-treated/treated split against the announcement date. Markets without a documented pre-announcement (Breathe 10.0, Magnifier 10.0, Radio 11.1, and Walkie Talkie 12.0) retain their release-date anchor. The *donut* specification holds the release-date anchor fixed but drops event-time months $\{-3, -2, -1\}$ from the regression sample and re-anchors the leads/lags reference period to $t = -4$; under this design, even if developers reacted in the three-month announcement window, that window is excluded from estimation regardless. The two specifications take different approaches to the same concern: the announcement shift recovers the effect if anticipation is the true treatment signal, while the donut simply drops the immediate-anticipation window without taking a stand on which date matters. A combined column applies both adjustments.

Table D.10 reports the four pooled ATTs for the seven primary incumbent outcomes at the headline market×time fixed-effects specification. The donut column is essentially identical to the baseline—Price moves from 0.422 to 0.428, Free from −0.058 to −0.061, Δ Log Rating Count from 0.017 to 0.022, and signs and significance hold on every outcome—so the immediate-anticipation window $\{-3, -2, -1\}$ is not driving the pooled estimates. The announcement-shifted column attenuates magnitudes on exactly the outcomes where we would expect three months of pre-release behavior to absorb some of the post-period response: Price falls from 0.42 to 0.30 (−29%), Free from −0.058 to −0.045 (−23%), and Δ Log Rating Count from 0.017 to 0.012 (−28%). Despite the attenuation, every signed effect retains the same direction and significance level as in the baseline column, with one exception: Price (Paid Apps), significant at the 10% level in the baseline and donut columns, is no longer significant once treatment is re-anchored to the announcement date.

The lone fragile estimate is the conditional price among paid apps, while the unconditional price effect and the decline in the free share remain strongly significant throughout. This reinforces the reading in Section 3.1.3 that the monetization response runs through the extensive margin, with incumbents switching from free to paid rather than paid apps raising prices. Average rating, if anything, strengthens modestly under the anticipation anchors (from 0.029 in the baseline to 0.034 under the announcement shift) and remains significant at the 1% level throughout. The combined column is close to the announced column, consistent with the donut adjustment being the smaller of the two. Together, the two specifications show that the headline TWFE estimates do not depend qualitatively on whether developers respond to the announcement date or the release date.

The same two specifications extend to the market-level entry/exit margin, where pre-entry dynamics are most visible in the event studies: the differential entry leads in Figure 4 rise to between +0.11 and +0.22 log points five to eight months before entry, then return fully to baseline by $t = -4$—an inverted U rather than a drift toward the treatment period. No individual lead is statistically distinguishable from zero, and there is no systematic platform-specific trend in pre-

Table D.10: Anticipation Robustness: WWDC-Shifted Treatment and Donut Specifications

Outcome	(1) Baseline	(2) Announced	(3) Donut	(4) Both
Δ Log Rating Count	0.0170***	0.0123***	0.0216***	0.0177***
	(0.0035)	(0.0035)	(0.0046)	(0.0046)
Free App	-0.0580***	-0.0447***	-0.0607***	-0.0471***
	(0.0092)	(0.0071)	(0.0096)	(0.0077)
In-App Purchases	-0.0061	-0.0035	-0.0058	-0.0023
	(0.0055)	(0.0054)	(0.0064)	(0.0062)
Price	0.4221***	0.2996***	0.4283***	0.3089***
	(0.0464)	(0.0337)	(0.0476)	(0.0368)
Price (Paid Apps)	0.0842*	0.0810	0.1008*	0.0736
	(0.0494)	(0.0519)	(0.0581)	(0.0635)
Avg. Rating	0.0293***	0.0344***	0.0360***	0.0396***
	(0.0069)	(0.0072)	(0.0082)	(0.0085)
Update	0.0104	0.0015	0.0107	0.0002
	(0.0077)	(0.0079)	(0.0089)	(0.0089)
N	59,616	58,788	49,680	48,990

Notes: TWFE pooled ATT estimates at the app-month level under four anticipation-robustness specifications. Column (1) is the baseline at the primary specification (cohort, market$\times$time fixed effects). Column (2) substitutes the earliest public Apple announcement date for the release date as the treatment anchor (18 of 22 active markets were pre-announced, median lead three months). Column (3) is a donut specification that drops event-time months $\{-3, -2, -1\}$ from the regression sample and re-anchors the leads/lags reference period to $t = -4$. Column (4) applies both adjustments. For the IAP outcome, 5 markets are excluded due to insufficient Android in-app purchase data in the pre-period. Standard errors clustered at the app level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table D.11: Entry/Exit Anticipation Robustness: WWDC-Shifted Treatment and Donut Specifications

Outcome	(1) Baseline	(2) Announced	(3) Donut	(4) Combined
Log Entry Count	-0.2487***	-0.2443***	-0.2790***	-0.2982***
	(0.0756)	(0.0743)	(0.0883)	(0.0729)
Log Exit Count	0.0167	-0.0223	0.0032	-0.0727
	(0.0718)	(0.0767)	(0.0768)	(0.0941)
N	792	792	660	660

Notes: TWFE pooled ATT estimates at the market-platform-month level under four anticipation-robustness specifications. Column (1) is the baseline at the primary entry/exit specification (market-platform and time fixed effects). Column (2) substitutes the earliest public Apple announcement date for the release date as the treatment anchor (18 of 22 active markets were pre-announced, median lead three months); the remaining markets retain the release date as the treatment anchor. Column (3) is a donut specification that drops event-time months $\{-3, -2, -1\}$ from the regression sample and re-anchors the leads/lags reference period to $t = -4$. Column (4) applies both adjustments. Standard errors clustered at the market-platform level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

entry entry counts: taking each market's pre-entry trend to be the OLS slope of log entry counts on event time over months $t = -9$ through -1, the mean within-market difference between the App Store and Play Store trends is -0.0078 log points per month, with a standard error of 0.0238. Table D.11 reports the results. The entry-deterrence estimate is stable under the announcement anchor (-0.244 against a baseline of -0.249) and, if anything, strengthens under the donut (-0.279) and combined (-0.298) specifications, remaining significant at the 1% level in all four columns. The post-entry decline is a level shift rather than the continuation of pre-entry dynamics: eight of the nine post-entry differential coefficients lie below the lowest pre-entry coefficient. The exit effect remains statistically indistinguishable from zero throughout.

D.7 Pre-Treatment Sample-Restriction Robustness

The incumbent sample restricts to apps that ranked in the top 200 for at least 10% of their days on the platform (Section 2.4, henceforth *200-p10*). The 10% requirement is evaluated over each app's entire observation window, including months after Apple's market entry, so the post-entry portion of an app's ranking history can affect whether the app qualifies. This subsection examines whether incumbent ATTs are sensitive to that mechanism by re-estimating the pooled TWFE specification under two alternative restrictions that bound the resulting selection on opposite sides.

The first alternative, *200-pt10*, applies the same 10%-of-operating-days top-200 rule but evaluates it only on rankings observed strictly before Apple's entry into each app's market, using the per-market entry dates from the registry. By construction *200-pt10* cannot be contaminated by

post-treatment outcomes. It is a strict subset of *200-p10*, so the difference between those two columns reflects apps whose top-200 status was acquired only after Apple's entry. The second alternative, *200-e1*, retains every app that ever ranked in the top 200 for at least one day, dropping the duration requirement entirely; this is the same rule used for the entry/exit and entering-cohort analyses (Section 3.1). Applied to incumbents, *200-e1* is approximately a superset of *200-p10*, so the difference between those columns reflects apps whose top-200 status *faded* after Apple's entry.

Table D.12 reports pooled TWFE incumbent ATTs at all three restrictions, holding fixed the market definition, fixed-effects structure, and inference approach used in the main text. Signs and significance are stable across columns for three of the five outcomes (Price, Avg. Rating, and Δ Log Rating Count). The exceptions are Update, insignificant under *200-p10* but significant under both alternatives, and the near-zero IAP estimate, whose sign flips within a narrow band around zero. Point estimates for Δ Log Rating Count, Avg. Rating, and Update sit within tight bands across the three samples, and the IAP estimate moves within a small range around zero. The Price effect is essentially unchanged moving from *200-p10* ($0.42) to *200-pt10* ($0.45) and shrinks to $0.08 under the much-larger *200-e1*. *200-e1* admits a large pool of marginally relevant apps that never sustained top-200 status; pricing responses among those apps are mechanically smaller, which attenuates the pooled estimate without reversing its sign. The qualitative conclusions of the incumbent analysis hold under all three restrictions.

The bias-channel rows in the table footer size both directions of the post-treatment selection. The first pair reports the share of *200-p10* that the pre-treatment-only rule drops because the app's top-200 status was acquired only after Apple's entry: 15.8% on iOS and 14.2% on Android. The near-equality of these rates indicates that *200-p10* does not differentially admit "post-entry success" apps across platforms *in volume*: whatever apps the rule pulls in from this channel, it pulls in at similar rates on iOS and Android. Equal rates speak to how *many* apps the rule selects, not to their characteristics or outcome trajectories. The more direct check on whether this selection contaminates the iOS-versus-Android comparison is that the ATT itself barely moves across restrictions, as shown by the small pooled shifts below and the market-level ATT correlations that follow. The second pair reports the share of *200-e1* that *200-p10* excludes because the app failed to sustain top-200 status across the panel: 80.1% on iOS versus 68.0% on Android. iOS apps in this broader pool are roughly 12 percentage points more likely to fall short of *200-p10*'s sustained-relevance window, a pattern consistent with some incumbent apps losing ranking traction after Apple's entry. This asymmetry sets a ceiling on the bias *200-p10*'s exclusion channel could induce. Four of the five outcomes shift by less than one pooled standard error between *200-p10* and *200-e1*, so the realized bias for those outcomes is far below what the channel size could in principle support; Price is the only outcome that moves materially.

The market-level decomposition that underlies Section 3.2 shows the same pattern more directly. Between *200-p10* and *200-pt10*—the comparison that isolates post-treatment selection—the market-level ATTs correspond almost exactly: the correlation exceeds 0.94 for every outcome and reaches 0.998 for Price, and no market with a statistically significant effect changes sign. The much

Table D.12: Sample-Restriction Robustness (Incumbent Apps): TWFE Pooled ATTs Across Three Restrictions

Outcome	(1) 200-$p10$ (Primary)	(2) 200-$pt10$ (Pre-treatment)	(3) 200-$e1$ (Less restrictive)
Price	0.4221***	0.4529***	0.0812***
	(0.0464)	(0.0530)	(0.0257)
In-App Purchases	-0.0061	0.0025	-0.0034
	(0.0055)	(0.0047)	(0.0025)
Avg. Rating	0.0293***	0.0277***	0.0294***
	(0.0069)	(0.0069)	(0.0042)
Update	0.0104	0.0148*	0.0097***
	(0.0077)	(0.0081)	(0.0035)
Δ Log Rating Count	0.0170***	0.0149***	0.0184***
	(0.0035)	(0.0036)	(0.0017)
# Apps	3,312	2,805	14,766
# Markets	22	22	22
# iOS app-months	41,976	35,352	210,582
# Android app-months	17,640	15,138	55,206
Bias-channel sizes (apps each alternative spec adds vs. drops from the primary *200-p10* sample, by platform; rate parenthetical)			
iOS apps in 200-$p10 \setminus$ 200-$pt10$ (post-entry success)	—	368 (15.8%)	—
Android apps in 200-$p10 \setminus$ 200-$pt10$	—	139 (14.2%)	—
iOS apps in 200-$e1 \setminus$ 200-$p10$ (post-entry decline)	—	—	9,367 (80.1%)
Android apps in 200-$e1 \setminus$ 200-$p10$	—	—	2,087 (68.0%)

Notes: TWFE pooled ATTs for incumbent app outcomes under three sample restrictions. All columns use app and market×month fixed effects, base matching, and the $\theta = 0.6$ market definition. The *200-pt10* column applies the *200-p10* rule (top-200 for ≥10% of operating days) only to each app's pre-Apple-entry window, so it is a strict subset of *200-p10*; the *200-e1* column requires only that an app ever ranked top-200 for ≥1 day (matches the cohort/EE panels). Standard errors clustered at the app level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The bias-channel rows report how many apps each alternative restriction drops from or adds to the primary *200-p10* sample, split by platform with the within-platform rate in parentheses (denominators: *200-p10* iOS/Android counts for the upper rows; *200-e1* iOS/Android counts for the lower rows). A large difference between the iOS and Android rates would indicate that *200-p10*'s reliance on whole-panel rankings shifts sample composition differentially across platforms around Apple's entry; comparable rates indicate that the iOS-vs-Android comparison sample is not differentially affected.

broader *200-e1* sample corresponds less closely, with correlations of 0.69 for Price, 0.80 for Δ Log Rating Count, and 0.68 for Avg. Rating. There the Camera market reverses from a \$1.22 increase to a \$0.79 decline—both significant at the 1% level—as the sample expands to every app that ever ranked in the top 200. That reversal, in the market that most drives the pooled Price effect, is the same dilution that pulls the pooled estimate down to \$0.08; it compresses the cross-market spread as well, with the standard deviation of market-level effects 36% to 62% lower under *200-e1* for every outcome. Because the pre-treatment-only restriction that directly addresses the post-treatment-selection concern leaves the largest pricing and quality effects unchanged, the heterogeneity story documented in the main text is not an artifact of post-treatment selection.

E Single-Homing Sample Analysis

A potential concern with the cross-platform identification strategy is that multi-homing apps (those available on both iOS and Android) may transmit treatment effects across platforms. If an App Store developer adjusts its strategy in response to Apple's entry, the Play Store version of the same app may also change, contaminating the control group. This appendix restricts the incumbent

sample to single-homing apps (those available on only one platform) and shows that the main findings are not driven by such spillovers.

E.1 Cross-Platform Record Linkage

Since no common identifier links apps across the App Store and Play Store, I identify multi-homing apps through probabilistic record linkage using the `recordlinkage` Python library. I generate candidate pairs via a Sorted Neighbourhood indexer on app names (window size 21), then compute Jaro–Winkler string similarity on three fields: app name, developer name, and developer URL. A pair is classified as a match if at least two of the three fields exceed a similarity threshold of 0.85. Apps with no cross-platform match are classified as single-homing. The record linkage is performed over the full population of apps observed on each platform, not the restricted analysis sample, so an app is not classified as single-homing merely because its cross-platform counterpart falls outside the incumbent sample. A counterpart that never appears in the data at all cannot be matched, which would, if anything, understate cross-platform overlap and leave the single-homing restriction retaining some apps that are in fact multi-homing.

E.2 Sample Characteristics

Table E.1 presents summary statistics for the single-homing sample. Restricting to single-homing apps reduces the incumbent sample from 3,312 to 1,915 unique apps (a 42% reduction). Panels B and C are unchanged, as market-level entry/exit and cohort composition analyses do not condition on individual app homing status.

E.3 Incumbent App Results

Table E.2 compares TWFE ATT estimates from the full sample and the single-homing sample. All seven outcomes have the same sign in both samples, and four of seven are statistically significant in the single-homing sample despite its smaller size. Monetization effects are, if anything, somewhat larger in the single-homing sample (the price increase rises from \$0.42 to \$0.44, and the decline in the share of free apps from 5.8 to 6.8 percentage points), consistent with spillovers modestly attenuating the full-sample estimates. Rating is significant in both samples, with a single-homing point estimate (0.025) close to the full-sample value (0.029). The one outcome that loses significance is the conditional price among paid apps, which falls from 0.084 to 0.046 and is no longer distinguishable from zero in the reduced sample; in-app purchases and update frequency are insignificant in both. As in the anticipation checks, this pattern, larger unconditional price and free-share effects alongside a weaker conditional price, reinforces the Section 3.1.3 reading that the monetization response operates on the extensive margin, with incumbents moving from free to paid rather than paid apps raising prices. The qualitative conclusions are unchanged.

Figures E.1 and E.2 present event studies for the single-homing sample. The price, free-share, in-app-purchase, and update leads are flat, while the average-rating and rating-count-growth leads

Table E.1: Summary Statistics (Single-Homing Sample)

	Pre-Treatment	Post-Treatment	Change
Panel A: Incumbent Apps			
Price	0.747	1.629	+0.882
Price (Paid Apps)	3.800	3.997	+0.197
Free	0.803	0.592	-0.211
In-App Purchases	0.350	0.369	+0.018
Avg. Rating	3.974	3.977	+0.003
Update	0.206	0.196	-0.010
Δ Log # Ratings	0.046	0.027	-0.019
N	17,235	17,235	
Panel B: Market-Level Entry and Exit			
Log Entry	1.499	1.394	-0.105
Log Exit	1.250	1.373	+0.123
N	396	396	
Panel C: Entering Cohorts			
Price	0.372	0.682	+0.309
Price (Paid Apps)	3.152	3.114	-0.038
Free	0.882	0.781	-0.101
In-App Purchases	0.143	0.163	+0.021
Avg. Rating	4.332	4.315	-0.017
Log # Ratings	0.796	0.955	+0.159
Similarity	0.648	0.650	+0.001
N	2,302	1,960	

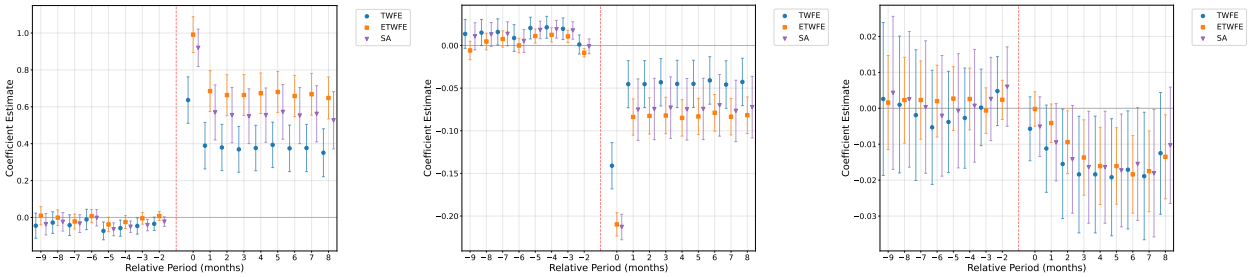
Notes: Panel A reports app-month level statistics for incumbent apps observed in an 18-month balanced panel around each entry event. Panel B reports market-platform-month level log entry and exit counts. Panel C reports statistics for apps observed once upon entry (repeated cross-section). Pre-Treatment refers to the 9 months before Apple's entry; Post-Treatment refers to the 9 months after.

Table E.2: Incumbent Treatment Effects: All Apps vs. Single-Homing Apps

Outcome	All Apps (1)	Single-Homing (2)
Monetization		
Price	0.422***	0.443***
	(0.046)	(0.064)
Free App	-0.058***	-0.068***
	(0.009)	(0.014)
In-App Purchases	-0.006	-0.015
	(0.005)	(0.009)
Price (Paid Apps)	0.084*	0.046
	(0.049)	(0.083)
Quality		
Avg. Rating	0.029***	0.025***
	(0.007)	(0.009)
Update	0.010	0.006
	(0.008)	(0.011)
Δ Log Rating Count	0.017***	0.029***
	(0.004)	(0.006)
Observations	59,616	34,470

Notes: This table compares TWFE ATT estimates using all apps (column 1) versus only single-homing apps (column 2). Single-homing apps are those available on only one platform, excluding apps identified as operating on both iOS and Android via record linkage. All specifications include app and market×month fixed effects with standard errors clustered at the app level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

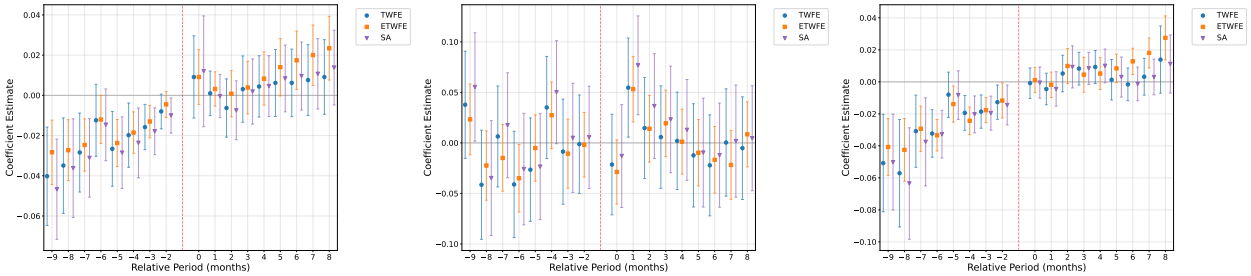
Figure E.1: Single-Homing Sample: Monetization Dynamics



(a) Price (All Apps) (b) Free (c) In-App Purchases

Note: Event study estimates for monetization outcomes in the single-homing sample. TWFE, ETWFE (Wooldridge, 2025), and SA (Sun and Abraham, 2021) are shown. All specifications use app and market×month fixed effects (except ETWFE, which uses app and month fixed effects). For the IAP outcome, 5 markets are excluded due to insufficient Play Store in-app purchase data in the pre-period. Standard errors are clustered at the app level. Error bars represent 95% confidence intervals.

Figure E.2: Single-Homing Sample: Quality Dynamics



(a) Rating (b) Update (c) Δ Log Rating Count

Note: Event study estimates for quality outcomes in the single-homing sample. TWFE, ETWFE (Wooldridge, 2025), and SA (Sun and Abraham, 2021) are shown. All specifications use app and market×month fixed effects (except ETWFE, which uses app and month fixed effects). Standard errors are clustered at the app level. Error bars represent 95% confidence intervals.

drift upward over the pre-period—the same pattern present in the full-sample event studies, not an artifact of the single-homing restriction. Both the leads and the post-entry dynamics track the main-text estimates closely.

E.4 Cross-Market Heterogeneity

Table E.3 reports market-level heterogeneity diagnostics for the single-homing sample. The I^2 statistics are comparable to the full-sample results in Section 3.2. As in the full sample, most markets are null and the few significant markets almost all confirm the pooled sign; the null share rises most for in-app purchases (from 81% to 100%), consistent with noisier market-level estimates from smaller per-market samples.

Table E.3: Cross-Market Heterogeneity in Incumbent Treatment Effects (Single-Homing Sample)

Outcome	Est. (SE)	% Conf.	% Opp.	% Null	I^2 (Q p)	$\sqrt{\tau^2}$
Monetization						
Price	0.443	27%	0%	73%	88%	0.200
	(0.064)				(<0.001)	
Free	-0.068	15%	0%	85%	80%	0.028
	(0.014)				(<0.001)	
In-App Purchases	-0.015	0%	0%	100%	46%	0.009
	(0.009)				(0.029)	
Price (Paid Apps)	0.046	0%	0%	100%	0%	0.000
	(0.083)				(0.749)	
Quality						
Rating	0.025	10%	0%	90%	3%	0.006
	(0.009)				(0.423)	
Update Frequency	0.006	5%	0%	95%	42%	0.037
	(0.011)				(0.025)	
Δ Log Rating Count	0.029	20%	0%	80%	46%	0.018
	(0.006)				(0.013)	

Notes: This table reports heterogeneity diagnostics applied to market-level ATTs. Est. is the TWFE estimate. For each outcome, markets are classified by the significance and direction of their market-specific effect: % Conf. is the share significant at the 5% level with the same sign as the pooled estimate, % Opp. the share significant with the opposite sign, and % Null the share not significantly different from zero. I^2 measures the proportion of total variation due to between-market heterogeneity rather than sampling error; Q p-value tests the null of homogeneous effects. $\sqrt{\tau^2}$ is the DerSimonian–Laird estimate of the between-market standard deviation of true effects.