# Commoditize Your Complement: Strategic Open-Source Releases and the Cannibalization of Rival Firms

Peter Young Won *

May 2026

**ABSTRACT**

Why do firms release open-source software (OSS) for free? I show that firms strategically release OSS to commoditize complementary product markets and erode rivals' proprietary rents. Using the staggered release of GitHub repositories, I find that the equity market prices each release as a directed wealth transfer from complement rivals to the releasing platform: within a three-day window the platform gains $313 billion while complement rivals lose $153 billion, a 2.8% value-weighted decline. This displacement is concentrated among complementary peers, while direct substitutes show no measurable reaction. Exposed rivals also lose 5.2% of sales, cut R&D by 14.8%, and reallocate engineering talent to the platform. Free OSS emerges as a non-priced competitive weapon outside the reach of traditional antitrust.

JEL Classification: L12, L17, O31, O32, G14, G32

Keywords: Open-Source Software, Commoditization, GitHub Ecosystem, Innovation

*Won: peter.young.won@eccles.utah.edu.

# 1 Introduction

Why do firms release open-source software for free? Over the past decade, Alphabet, Meta, Microsoft, and a handful of other dominant technology firms have published billions of lines of source code to the public at no charge. Google gave away TensorFlow, the framework that runs its own search and advertising models, and Meta did the same with PyTorch and React. Each release hands rivals software at no cost that took years and hundreds of engineers to build. A growing literature treats these releases as a public good that raises productivity for the firms that adopt them (Nagle, 2019; Hoffmann et al., 2024). Platform theory points to a second possibility: a multi-segment firm can rationally subsidize a complement when doing so commoditizes a layer of the stack on which its rivals depend, raising demand for its own protected core (Parker and Van Alstyne, 2005; Economides and Katsamakas, 2006; Gawer and Henderson, 2007). Whether this commoditize-your-complement mechanism redistributes rents away from incumbents rather than lifting all firms together has not been measured. I provide the first large-sample evidence.

A historical precedent makes the mechanism concrete. Gurley (2026) observes that the surprising outcome of cloud computing's first two decades is not that Amazon Web Services came to dominate, but that IBM, Hewlett-Packard, and Intel, the owners of the underlying stack, did not: open-source commoditization of that stack dissolved their supplier power, and the operator captured the rents instead. This paper documents the same dynamic arising by deliberate choice, with dominant incumbents commoditizing complementary layers through their own releases.

Measuring the competitive effect on incumbent rivals requires three ingredients at once: a firm-level measure of exposure to each release, a set of release shocks comparable in economic magnitude, and a credible counterfactual for the same firms absent the release. The existing literature on open source supplies none of the three, having concentrated on the releasing firm or on aggregate productivity (Nagle, 2019; Hoffmann et al., 2024). I assemble all three. I combine repository-level developer activity from GitHub Archive with firm-level Compustat outcomes and USPTO patent records, I identify roughly 1,700 economically consequential releases with a market-value intensity measure adapted from Kogan et al. (2017), and I map each release to its product-market peers using text-based network industry classifications (Hoberg et al., 2014; Hoberg and Phillips, 2016). A stacked difference-in-differences estimator, weighted by ex-ante release intensity, then traces how each cohort reshapes peers' revenues, costs, innovation, and labor.

The rival response unfolds in three steps. First, I find that revenue contracts: treated peers' sales fall by 5.2 percent relative to matched controls in the post-release window. Facing this compression, the same firms cut direct operating costs by substituting the now-free open-source infrastructure for internal proprietary inputs, which produces a short-run lift in operating margin that masks the underlying erosion.

Second, I show that this cost adjustment carries a long-run innovation cost. Research and development scaled by assets falls by 14.8 percent over the same window, the average age of treated firms' patent portfolios rises, and the share of frontier exploratory patents declines. The retreat shows up in product-market fluidity, which contracts: exposed incumbents do not rotate into adjacent product spaces but let their offerings stagnate.

Third, I find that the engineering labor market links the cost cuts to the innovation decline. Using developer engagement logs from GitHub for the most strategically influential releases, I observe that rival engineering capital flows onto the focal platform's projects: active development staffing at exposed peers rises by about 30 percent and active code contributions by about 50 percent. Proprietary innovation contracts at the same time that competitor engineers spend a growing share of their effort on the focal firm's commons (Gupta et al., 2024).

These averages conceal a sharp asymmetry. I partition treated peers by Cooperative Patent Classification overlap and announcement-window return co-movement into substitute peers, which compete head to head with the focal firm, and complement peers, which operate one layer adjacent to it. I find that the revenue, cost, innovation, and labor effects load almost entirely on complement peers, while substitutes are statistically indistinguishable from zero on every margin. The equity market prices the same split in real time. In the three-day window around release, value-weighted cumulative abnormal returns are $+0.5$ percent for releasing platforms, $-2.8$ percent for complement peers, and zero for substitutes, a wealth transfer that aggregates to \$313 billion gained by releasing platforms and \$153 billion lost by complement rivals. The product-market and asset-price evidence line up on the same firms, which rejects diffuse spillover and identifies the release as a targeted shock to the platform's complementary layer.

The identification rests on three design choices. Peer assignment is anchored to the lagged TNIC product-market network, so a release cannot rewrite the peer set it is supposed to affect.

Cohorts are weighted by ex-ante Kogan release intensity, so the estimate reflects the economically influential events rather than the long tail of trivial ones. Controls are one-to-one propensity-score matches drawn from outside the focal firm's network (Baker et al., 2022). Under this design, the post-release dynamics are not pre-existing industry trends but the direct consequence of the release.

This paper makes three contributions. First, it provides the first large-sample, rival-level test of the commoditize-your-complement logic (Parker and Van Alstyne, 2005; Economides and Katsamakas, 2006; Gawer and Henderson, 2007). The empirical literature on open source has measured aggregate productivity gains (Nagle, 2019), the financing of complementary startups (Conti et al., 2025), and the private value captured by the contributing firm (Hoffmann et al., 2024; Zhao, 2025), but the rival whose proprietary market is commoditized has not been the unit of analysis. I locate the strategy at the rival level, I translate the market reaction into dollar-valued wealth transfers, and I show that a continuous text-based exposure measure separates the harmed peers from the insulated ones.

Second, I extend the resource-based view and the open-innovation literature to the involuntary case. That literature argues that voluntary surrender of proprietary control can be rational for the disclosing firm (Alexy et al., 2018, 2013; Chesbrough and Crowther, 2006). I document the other side of the transaction: when a dominant platform releases a complement at zero price, exposed rivals not only watch the value of their proprietary stack erode but also lose engineering talent to the platform's commons. Labor reallocation and innovation decay therefore emerge as the hidden costs of defensive open-source adoption (Mkrtchyan et al., 2025; Heath et al., 2023; Gupta et al., 2024), widening the set of margins along which competitive harm travels in digital product markets.

Third, the findings speak to a regulatory gap. Traditional antitrust screens platform conduct for predatory pricing, output restriction, or explicit exclusion in the core market. A free release of a complementary technology satisfies none of these doctrines, yet I show that it extracts rents from rivals, slows their independent innovation, and absorbs their engineering capacity. As data and digital infrastructure increasingly function as networked competitive assets (Bian et al., 2025), an antitrust framework that price-screens only the focal product will miss the non-priced competitive instrument documented here.

The remainder of the paper is organized as follows. Section 2 positions the paper relative to five strands of prior work. Section 3 develops a multi-segment platform model that delivers the asymmetric sign prediction tested in the main analysis. Section 4 describes the data and the stacked difference-in-differences identification strategy. Section 5 reports the baseline rival-side findings on revenue, cost, labor, innovation, and product-market dynamics. Section 6 introduces the substitute–complement partition, shows that the average effects load on complement peers, and closes with the contemporaneous wealth-transfer evidence from cumulative abnormal returns. Section 7 concludes.

## 2 Related Literature

This paper draws on five lines of prior research: the theory of platform strategy and the commoditize-your-complement logic, empirical work on strategic open-source release and selective revealing, the economics of information goods and two-sided markets, the measurement of competitive rivalry and technology spillover, and the resource-based view (RBV) of the firm. Each supplies part of the vocabulary for a commoditize-your-complement strategy, but none predicts the sign asymmetry this paper documents, in which substitute peers are insulated while complement peers are harmed. Establishing that asymmetry as an equilibrium prediction rather than an ex-post heterogeneity result is the paper's main theoretical contribution.

The idea that a firm might accept lower rents in one market to raise rents in another, articulated as industry folklore by Spolsky (2002), was formalized in platform economics: Parker and Van Alstyne (2005) show that an information good with strong cross-network externalities on a paid core service is optimally priced at zero, Economides and Katsamakas (2006) extend this to the proprietary-versus-open-source regime choice, and Gawer and Henderson (2007) describes the managerial practice of cultivating complementary ecosystems. A parallel strand studies what happens to rivals when a platform acts aggressively, including Eisenmann et al. (2011) on platform envelopment and Wen and Zhu (2019), Cennamo et al. (2018), and Wen et al. (2016) on how the threat of platform-owner entry reshapes complementor behavior and investment incentives. These frameworks explain why a profit-maximizing firm subsidizes a complement, but they treat the affected rivals as undifferentiated, and they typically have the focal firm enter the adjacent market, whereas an open-source release commoditizes that market without entry. The closest structural antecedent of this paper's peer classification is Adner and Kapoor (2010), who distinguish upstream components from downstream complements and show that performance across technology generations depends on which side of that divide a firm occupies; this paper operationalizes that distinction as the Substitute–Complement peer partition and shows it is, empirically, the primary axis along which release effects sort.

A more empirically grounded literature documents the strategic motives behind open-source release at the firm and project level: Lerner and Tirole (2002) on why firms contribute at all, Henkel (2006) on selective revealing in embedded Linux, Fosfuri et al. (2008) on commercializing releases as a product-market strategy, Alexy et al. (2013) and Alexy et al. (2018) on when surrendering proprietary control benefits the revealing firm, and Casadesus-Masanell and Llanes (2011) on profitable mixed open-and-proprietary strategies. A recent wave uses firm-level data to estimate the private value of adopting or contributing to open source (Nagle, 2019; Hoffmann et al., 2024; Zhao, 2025; Nagaraj and Piezunka, 2024; Haese and Peukert, 2025) and how participation interacts with corporate investment and financing (Conti et al., 2025; Emery et al., 2025; Coleman et al., 2025). This work has concentrated on the releasing or adopting firm and has relied largely on cross-sectional adoption variation or case studies; the rival whose proprietary business is commoditized has rarely been the unit of analysis. This paper instead treats approximately 1,700 economically consequential releases as staggered shocks and places the rival firm's financial, innovation, and labor outcomes at the center of measurement.

The pricing logic behind free complements is developed in the two-sided-markets and network-effects literatures: Rochet and Tirole (2003) and Armstrong (2006) on optimally subsidizing one side to expand the other, Katz and Shapiro (1985) and Farrell and Saloner (1985) on how network and standardization benefits make zero or even negative prices equilibrium-consistent, and Shapiro and Varian (1999) on how zero marginal cost combined with consumer lock-in makes free release a potentially dominant strategy for digital information goods. The closest empirical antecedent, Casadesus-Masanell and Ghemawat (2006), models an open-source entrant coexisting with a proprietary incumbent under strong demand-side learning, but treats the open-source good as arising bottom-up from a community rather than as a top-down release by a multi-segment incumbent that internalizes both the rent loss in the commoditized segment and the rent gain in the core segment. That internalization is what turns rival displacement into an equilibrium outcome rather than a byproduct of community competition.

A large literature uses firm-level data to separate the two forces linking a firm's innovation to its rivals, positive knowledge spillover and negative product-market rivalry: Jaffe (1986) established the measurement approach and Bloom et al. (2013) sharpened identification and found that, in aggregate, spillovers dominate rivalry, while Aghion et al. (2005) document an inverted-U between competition and innovation and Makaew and Maksimovic (2020) and Baker et al. (2023) trace competitive shocks through supplier–customer links. This paper is the mirror image of Bloom et al. (2013): the release is itself the spillover, and the focal firm chooses its scope and timing, so the net rival effect does not average to a single sign but splits, harming substitutes in the commoditized market while complements in adjacent markets benefit from ecosystem expansion, a split that is invisible in a specification treating spillover as passive. Seamans and Zhu (2014) on Craigslist's effect on newspaper classified revenue and Boudreau (2010) on how platform openness shapes downstream innovation show that a single platform shock can produce large, identifiable rival effects; this paper extends both across a portfolio of roughly 1,700 releases and shows the rival response is partitioned by product-market position. Finally, the resource-based view treats competitive advantage as rooted in valuable, rare, and inimitable resources (Barney, 1991) and has been extended to voluntary openness (Alexy et al., 2018; Chesbrough and Crowther, 2006), but has paid less attention to the involuntary case in which a third party's strategic release erodes the rarity and inimitability of a rival's resource stock and forces adjustments along its innovation and labor-allocation margins (Benner and Tushman, 2002; Manso, 2011; Mkrtchyan et al., 2025; Lin and Maruping, 2022; Heath et al., 2023; Gupta et al., 2024).

Taken together, these literatures supply the theoretical vocabulary but not the empirical architecture to identify a commoditize-your-complement strategy as it operates on rivals: platform strategy explains the focal incentive but treats rivals as undifferentiated; the strategic-OSS literature identifies the releasing firm's motives but rarely measures the rival's response; the two-sided-markets and standards literatures explain the zero price but not which rivals lose; the spillover literature documents aggregate rival effects but treats spillover as passive; and the RBV explains which resources are vulnerable but not the mechanism of third-party erosion. This paper combines an identified large-sample event design, a rival-level outcome panel, and a Substitute–Complement partition grounded in the upstream–downstream distinction of Adner and Kapoor (2010), and traces the economic tradeoff in which a focal firm gains in its profit-center segment by commoditizing a complementary one, transferring wealth from rivals whose revenues reside in that segment while leaving rivals in adjacent segments untouched or even better off. No single line above, taken on its own, generates this prediction.

# 3 Theoretical Framework and Hypothesis Development

This section sketches the framework that organizes the empirical analysis and states the six hypotheses it generates. The mechanism, in the vocabulary of Porter's Five Forces, is the elimination of upstream supplier power through commoditization of the complement layer, with the rent then accruing to the platform operator rather than the IP owner. Gurley (2026) describes Amazon Web Services as the canonical historical instance: exogenous open-source commoditization of the operating-system and middleware stack left the operator position empty for Amazon to occupy, while IBM, Hewlett-Packard, and Intel, the firms that should have extracted the supplier rents, were left with no surface on which to charge. The framework below makes the same logic endogenous, with incumbents financing the commoditization themselves out of profit-center rents. The formal model, comprising the multi-segment profit function, the four-channel decomposition of the release decision, and the derivations of Propositions 1 and 2, is developed in full in Appendix A.

The analytical primitive is the multi-segment firm, operating across three logically distinct product markets. Segment $A$ is the profit center that generates the firm's protected rents, such as advertising inventory for Meta, search advertising for Alphabet, and enterprise cloud for Microsoft. Segment $B$ is a substitute market in which rival entry threatens the $A$-segment boundary. Segment $C$ is a complement market in which higher consumption raises the marginal value of $A$ through a cross-market demand linkage. In segment $C$ the firm's optimal posture is commoditization rather than rent extraction, because cheaper $C$-goods expand demand for the profitable $A$-segment. A strategic open-source release is exactly this act: the focal firm gives away a high-quality asset in segment $C$ at a consumer price of zero in perpetuity, financed out of segment $A$'s rents. The open-source license makes the giveaway both irreversible and cross-subsidized, which is what separates it from an ordinary price cut and renders the standard predatory-pricing hold-up logic inapplicable.

Releasing is optimal when the cross-market subsidy to segment $A$, together with the option value of setting the technical standard and of co-opting outside developer effort, dominates the direct loss of segment-$C$ rent. This condition holds for incumbents whose profit base is dominated by a large, protected $A$-segment, the firms with a subsidy pool deep enough to finance a permanent giveaway, and fails for firms whose profits are themselves concentrated in segment $C$. Because a release with negative net present value would not be announced, the focal firm's announcement-window equity return is non-negative in expectation.

**Proposition 1 (Release under Profit-Center Dominance):** *A focal firm releases a strategic open-source asset in its complement market when its profit-center rents are large relative to its complement-segment rents and the cross-market demand linkage is sufficiently strong. The release is then optimal even though it forgoes the firm's own complement-segment revenue.*

**Hypothesis 1 (Focal-Firm Valuation Gain):** *A strategic open-source release generates a non-negative cumulative abnormal return on the focal firm's equity in a narrow window around the release.*

The framework's sharpest prediction concerns the cross-section of peer firms, and it is pinned down by where each peer's revenue lives across the three segments. A *substitute peer* earns its revenue in the focal firm's profit center $A$; a *complement peer* earns it in segment $C$, the very market the release commoditizes. For a complement peer, the release directly compresses pricing power: consumers facing a free, high-quality alternative abandon its proprietary $C$-product, and its revenue base contracts with no offsetting revenue in $A$-adjacent markets. For a substitute peer, there is no direct $C$-segment compression, and the only effect is the modest, second-order benefit of an expanded complement market, which is weakly positive or statistically indistinguishable from zero.

**Proposition 2 (Asymmetric Response):** *Announcement of a strategic open-source release generates a strongly negative cumulative abnormal return on complement peers and a weakly positive or null return on substitute peers. The complement-peer loss is larger the more the peer's revenue is concentrated in the commoditized segment, while the substitute-peer effect is bounded above by the focal firm's own absorption of the expanded complement-market demand.*

**Hypothesis 2 (Asymmetric Peer Valuation):** *A strategic open-source release generates a strongly negative cumulative abnormal return on complement peers and a response statistically indistin-*

*guishable from zero on substitute peers; the asymmetry follows from peer portfolio composition and assumes nothing about the focal firm's targeting intent.*

Hypothesis 2 is the load-bearing prediction of the framework, and it is a discriminating one: no alternative account of the release, whether quality signaling, corporate philanthropy, Bertrand predation, or network bootstrapping, reproduces a strong negative response concentrated on complement peers together with a null on substitutes. Section 6.1 details how the latent substitute and complement types are projected onto observable pre-event signals.

The negative valuation response on complement peers propagates into firm behavior across four real-economy margins. Facing a credible, permanent commoditization of its $C$-segment product, the complement peer loses pricing power and its revenue contracts; at the same time it adopts the free release as an input substitute, cutting its cost of goods sold fast enough to expand operating margins in the short run. With the marginal return to proprietary $C$-innovation falling, internal exploratory research becomes a dominated strategy and the peer's patent portfolio ages. Its engineering labor is redirected toward maintaining compatibility with the focal firm's evolving standard, measurable as contribution activity on the focal firm's public repositories. The cumulative retreat freezes the peer's external product-market positioning.

**Hypothesis 3 (Financial Compression and Short-Run Margin Expansion):** *Complement peers exhibit declining sales and declining cost of goods sold, with the cost contraction expanding operating margins in the short run.*

**Hypothesis 4 (Innovation Retreat):** *Complement peers reduce internal R&D expenditure and experience aging of their exploratory patent portfolios.*

**Hypothesis 5 (Labor Co-optation):** *Complement peer developers increase active contribution (commits, pull requests, and issues) to the focal firm's open-source repositories.*

**Hypothesis 6 (Product-Market Freeze):** *Complement peers experience a decline in product-market fluidity.*

Hypotheses 3 through 6 map one-to-one onto the four channels of the release decision and together characterize the complement peer's full adjustment path. On the producer side the framework predicts a focal-firm gain, a complement-peer loss that exceeds it in dollar terms, and an approximate substitute-peer null; the net effect on consumer surplus is left open, and the paper documents the wealth transfer as a positive economic fact. Appendix A develops the formal model and all derivations.

# 4 Data and Sample Construction

This study constructs a novel, multi-source panel designed to isolate the strategic impact of open-source software releases by dominant technology platforms on the financial performance, innovation trajectories, and human-capital allocation of product-market rivals. The data construction integrates four independently sourced components. First, I collect repository-level metadata and release histories from the GitHub REST API.[1] Second, I extract daily developer-activity logs from the GH Archive on Google BigQuery.[2] Third, I incorporate firm-year accounting, market, and patent data from Compustat, CRSP, and the USPTO. Finally, I map the competitive landscape using text-based product-market network measures from the Hoberg-Phillips TNIC database.[3] The primary empirical contribution of this paper lies in the integration of these components. Rather than selecting a discretionary set of high-profile releases, I identify the economically consequential event set using a market-value-based intensity measure adapted from Kogan et al. (2017), utilizing this intensity as the analytic weight in the stacked difference-in-differences estimator throughout the analysis.

## 4.1 Identifying Strategic Open-Source Releases: The Kogan Intensity Measure

The construction of the event sample makes two distinct methodological choices, one about which releases enter the panel and one about how each included release is allowed to influence the estimate. Treating these as separate decisions, rather than collapsing them into a single ex-post

[1]The official documentation for the GitHub REST API is available at `https://docs.github.com/en/rest`.

[2]Historical GitHub event logs are accessed via the GH Archive public dataset hosted on Google BigQuery, detailed at `https://www.gharchive.org/`.

[3]The product market data are publicly available at the Hoberg-Phillips Data Library: `https://hobergphillips.tuck.dartmouth.edu/`.

selection rule, is what addresses the two competing concerns about the event sample: selection on outcome-correlated characteristics, and signal attenuation from economically marginal releases.

On the sample side, I depart from the prevailing practice in the corporate open-source literature of restricting the event set to a discretionary list of high-profile releases identified through cumulative community stars, fork counts, or downstream adoption. Each of those metrics is realized over years of community uptake and is therefore mechanically correlated with the same competitive outcomes the analysis seeks to estimate; using them as a selection filter embeds look-ahead bias with respect to the response the research design is intended to measure. To eliminate this circularity, I include every open-source repository released by one of the focal technology firms between January 2015 and December 2024 for which CRSP coverage of the releasing firm and pre-event market capitalization are available. The single inclusion filter is therefore mechanical and ex-ante, not curatorial, and the resulting universe of approximately 1,700 release cohorts is fixed before any outcome is measured.

On the weighting side, I assign to each release a continuous, market-implied intensity weight constructed entirely from pre-event and contemporaneous stock-market information. For repository $r$ released by firm $f(r)$, the cohort-level Kogan intensity is

$$\kappa_r \;=\; \mathrm{CAR}^{\mathrm{FF4}}_{r,[-1,+1]} \times \mathrm{MktCap}_{f(r),-1},$$

where $\mathrm{CAR}^{\mathrm{FF4}}_{r,[-1,+1]}$ is the cumulative abnormal return of $f(r)$ over a three-trading-day window centered on the release announcement, estimated under the Fama and French (1992) four-factor model with factor loadings fit on a 150-day window ending 30 days before the event, and $\mathrm{MktCap}_{f(r),-1}$ is the firm's market capitalization on the pre-event day. The construction directly parallels the patent-value measure in Kogan et al. (2017), in which the dollar-value impact of a patent grant is recovered from the issuing firm's narrow-window abnormal return multiplied by its pre-event market capitalization. Applying the same logic to open-source releases converts an unranked technological event into a continuous, market-implied dollar-magnitude proxy. I use the absolute value $|\kappa_r|$ as the analytic weight throughout the stacked difference-in-differences specification, since the focus is on the structural magnitude of the competitive shock rather than on the releasing firm's directional wealth gain or loss.

The two choices address two distinct concerns that the prior literature typically conflates. Fixing the included event set at the mechanical 1,700-cohort universe rules out selection on outcome-correlated characteristics: nothing about a release's eventual community success or downstream adoption gates its presence in the panel. Weighting by $|\kappa_r|$, in turn, rules out attenuation bias from economically marginal releases: each cohort is weighted by its ex-ante market-implied dollar magnitude rather than receiving equal influence in the estimate, so the headline coefficient reflects the response to the economically consequential events rather than the average of consequential and inconsequential ones. Internet Appendix B.1 confirms that these two concerns are empirically separated as the design intends. A saturation curve that re-estimates the headline coefficient on Top-$K$ subsets of cohorts ranked by $|\kappa_r|$ shows that the $|\kappa_r|$-weighted estimate is essentially invariant to the sample cut, while a complementary trimming-from-bottom exercise shows that the unweighted estimate exhibits the textbook attenuation pattern the weighting is designed to correct.

Figure 1 plots the annual distribution of the 1,700 release cohorts. Consequential releases concentrate between 2015 and 2017, the period when leading technology firms decisively pivoted toward open-source leadership, followed by a meaningful secondary wave in 2019 and 2020. This continued flow provides the staggered temporal variation needed for the stacked difference-in-differences estimator.

*[Insert Figure 1 Here]*

Table 1 reports descriptive evidence that the intensity measure captures meaningful cross-cohort variation in economic magnitude. Panel A documents a wide dispersion across $\kappa_r$ quartiles, ranging from an average loss above 16 billion dollars in the bottom quartile to a gain above 19 billion dollars in the top quartile. Average developer engagement measured by cumulative GitHub stars remains uniformly high across quartiles, indicating that all cohorts in the universe represent substantive technological events; the intensity weight differentiates them on market-implied dollar magnitude rather than on adoption per se. Panel B shows the extreme concentration of consequential deployment among a small number of dominant platforms: Microsoft, Alphabet, and Meta together account for the majority of repositories and of cumulative intensity, motivating the focus on those firms as the primary drivers of platform-level commoditization.

[Insert Table 1 Here]

Figure 2 plots the cohort-level distribution of $|\kappa_r|$. The raw distribution is heavily right-skewed, in line with standard distributional properties of innovation and patent-value measures; the log-transformed distribution closely approximates a normal curve, neither bimodal nor disproportionately driven by a small cluster of outliers. This continuous log-normal shape supports the use of $|\kappa_r|$ as a continuous analytic weight in the stacked difference-in-differences specification, ensuring that the empirical estimates are weighted by the underlying economic magnitude of the release rather than treating every repository as an equal shock.

[Insert Figure 2 Here]

Two additional features of the intensity measure are worth stating explicitly. First, $\kappa_r$ is computed entirely from pre-event and contemporaneous market data, ensuring that it remains uncontaminated by the subsequent product-market adjustments, patenting responses, or labor reallocations evaluated as outcomes in this study. Second, relative to the original Kogan et al. (2017) construction, this adaptation does not subject the release announcement window to an exhaustive news-filtering procedure, so a marginal fraction of the intensity values may absorb concurrent unrelated corporate news. The saturation and trimming exercises in Internet Appendix B.1 indicate that this residual noise exerts negligible influence on the headline estimate, consistent with the inverse-variance logic that motivates the weighting choice in the first place.

## 4.2 GitHub Repository and Developer Data

To construct the universe of open-source releases and observe developer-side labor reallocation at a daily frequency, I assemble a granular dataset utilizing the GitHub REST API and Google BigQuery. This collection and refinement pipeline is explicitly designed to isolate externally sourced engagement from internal corporate development.

I first identify the focal technology firms whose corporate accounts drive the modern open-source ecosystem, including large platform firms such as Microsoft, Alphabet, Meta, and Apple, alongside infrastructure and semiconductor firms such as Nvidia, Adobe, and Oracle. Focusing on this concentrated set of firms is economically motivated. The open-source landscape is heavily skewed toward the top, where foundational platforms, standards-setting libraries, and high-impact infrastructure projects are overwhelmingly authored by a relatively small set of large technology firms. I map each focal firm to its authenticated organizational accounts, such as google, tensorflow, and microsoft. To prevent the inclusion of unaffiliated mirrors or community forks, I verify organizational membership by matching developer email domains to the firm's verified corporate domains, such as microsoft.com and oracle.com. Following this verification, I extract metadata for all public repositories owned by these accounts released between January 2015 and December 2024. Applying the mechanical inclusion filter described in Section 4.1 yields the approximately 1,700 eligible cohorts utilized throughout the main analysis, each subsequently entering the panel under its $|\kappa_r|$ analytic weight. For the mechanism analysis of developer-level labor reallocation detailed in Section 5.3, I restrict the sample to a top 100 sub-panel with matched BigQuery-level developer activity. This restriction is necessary because BigQuery's cost-constrained throughput precludes exhaustive developer-level data extraction across the full sample of 1,700 cohorts.

Using the GH Archive on BigQuery, I extract all event-level activity records for the target repositories, including pushes, pull requests, issue submissions, stars, and watches. I classify these records into two economically distinct categories. Active contributions, comprising pushes, pull requests, and issues, capture direct developer effort and committed human capital. Passive contributions, consisting of stars and watches, capture broader community attention and technological adoption. This bifurcation maps directly onto the theoretical distinction between labor reallocation, which serves as the core mechanism of commoditization in this study, and adoption spillovers, which shape longer-run platform network effects.

Linking individual GitHub developers to their employing firms constitutes the central measurement challenge in the labor reallocation analysis. I extract email domains and self-reported company names from developer profiles and commit logs, matching them to the verified corporate domains of publicly traded firms. These matched domains subsequently resolve to standard Compustat GVKEY identifiers. To ensure identification accuracy, I strictly drop generic personal email providers, such as gmail.com and yahoo.com, from the matching process. This econometric choice deliberately privileges a low Type I error rate at the cost of a higher Type II error rate. While excluding generic domains fails to capture corporate developers contributing via personal accounts, it essentially eliminates the risk of misattributing independent or hobbyist developer activity to a specific public firm. Consequently, the resulting developer-to-firm panel represents a conservative lower bound on actual corporate engagement. Any statistically significant treatment effect documented in this analysis thus survives substantial measurement-error attenuation toward zero.

Finally, to guarantee that the measured engagement reflects genuine external labor reallocation rather than the focal firm's internal research and development, I impose a strict self-loop filter. All GitHub activity generated by developers affiliated with the focal releasing firm is entirely removed before aggregation. For instance, contributions by Microsoft employees to a Microsoft-owned repository are excluded from the peer engagement metrics. The filtered panel therefore cleanly captures the external labor mobilization of competing firms redirecting their human capital toward the newly established open-source infrastructure.

## 4.3 Firm Financials, Market Valuation, and Innovation Outcomes

To evaluate the financial, operational, and innovation consequences of an open-source release, I construct a firm-year panel using accounting data from Compustat, market data from the Center for Research in Security Prices (CRSP) via the CRSP-Compustat Merged database, and patent data from the United States Patent and Trademark Office. The sample spans fiscal years 2010 through 2024 to accommodate a symmetric pre- and post-event window around every eligible open-source release. Following standard practice in empirical corporate finance, I exclude financial institutions and regulated utilities, as their capital structures are not comparable to those of the technology and industrial firms that populate the peer groups in this study.

From this integrated panel, I compute the primary outcome variables utilized throughout the analysis. To evaluate revenue impact and profitability, I construct measures for Sales, Cost of Goods Sold (COGS), and Return on Assets (ROA), each scaled by total assets to adjust for firm size. To measure innovation outputs, I calculate Research and Development (R&D) expenditure scaled by total assets. Furthermore, following Manso (2011) and Balsmeier et al. (2017), I classify patents as either exploratory or exploitative using Cooperative Patent Classification (CPC) codes. Exploratory patents are defined as those filed in CPC classes new to the focal firm, whereas exploitative patents are those filed in the firm's incumbent classes. Based on this classification, I compute the average age of the firm's overall patent portfolio and the average age of its exploratory patents. Finally, the firm-level control variables utilized in the baseline regressions, specifically

Firm Size, Market-to-Book Ratio, and Leverage, are constructed following the conventions established by Fama and French (1992), Frank and Goyal (2009), and Roberts and Whited (2013).

## 4.4 Text-Based Product Market Networks and Fluidity

To identify product-market peers and measure competitive repositioning as an outcome, I utilize the Text-based Network Industry Classifications (TNIC) and associated product-space measures developed by Hoberg and Phillips (2016).[4] Unlike static industry codes, the TNIC methodology computes continuous pairwise similarity scores based on the cosine similarity of firms' 10-K product descriptions. This yields a dynamic map of competition that is particularly well-suited to the fast-evolving technology sector. I employ TNIC similarity scores to identify the product-market peers of each focal releasing firm. Furthermore, I utilize the Product Market Fluidity and Product Market Scope measures from Hoberg et al. (2014), alongside a Doc2Vec-based self-fluidity measure, as structural outcome variables to capture how peer firms reshape their product vocabulary following the competitive shock of a release.

## 4.5 Treatment Assignment, Matching, and Cohort Construction

For each eligible open-source release $r$ by a focal firm $f(r)$, treated peers are defined as all Compustat firms exhibiting strictly positive TNIC similarity to $f(r)$ in the fiscal year immediately preceding the release, $t-1$.[5] Utilizing the lagged product-space network deliberately eliminates the most severe source of look-ahead bias. If I utilized the contemporaneous year-$t$ network, the open-source release itself could force peer firms to rewrite their product descriptions, meaning the treatment assignment would endogenously reflect the outcome under study. Anchoring the network at $t-1$ forces the treatment assignment to be strictly measurable from information available prior to the event. Combined with the Kogan intensity metric, which likewise relies solely on pre-event and

[4]The data are available at the Hoberg-Phillips Data Library, `https://hobergphillips.tuck.dartmouth.edu/`.

[5]A small share of treated peers are themselves focal firms in other release events; for example, Oracle and IBM appear as peers in Microsoft's release windows, and vice versa. As a robustness check, I drop the 5% of (focal, peer, event) observations in which the peer firm is itself a focal in the sample. The Substitute-minus-Complement peer-CAR difference moves by less than two basis points (from $-13.4$ to $-13.6$ bp, both significant at $p < 0.005$), and the coefficients on focal CAR and Kogan intensity are essentially unchanged. Cross-event focal-as-peer contamination is therefore not driving the asymmetric peer response.

narrow-window market information, this approach yields an identification strategy based entirely on ex ante information. This ensures precise identification of both the specific firms that receive treatment and the intensity of the competitive shock they experience.

Peers operating within the TNIC network differ systematically from the broader Compustat universe along critical dimensions such as firm age, growth orientation, and R&D intensity. Consequently, an unadjusted comparison to all non-peer firms would conflate the treatment effect with selection bias based on firm characteristics. I therefore construct a matched control group via one-to-one nearest-neighbor propensity score matching estimated at $t-1$. The candidate control pool consists of all Compustat firms outside the focal firm's TNIC network in the pre-event year. Propensity scores are estimated using Firm Size, Market-to-Book Ratio, Return on Assets, and Leverage. Matching at $t-1$ ensures that control firms resemble the treated peers based on pre-treatment fundamentals rather than realized outcomes.[6]

The final stacked panel concatenates the cohort-level event panels around each of the approximately 1,700 eligible releases. All continuous financial variables are normalized by total assets or sales and winsorized at the 1st and 99th percentiles to mitigate the influence of outliers. Table 2 reports the pre-event year ($t=-1$) summary statistics for both the treated and matched-control panels.

[Insert Table 2 Here]

Fundamental firm characteristics, including Market Capitalization, Market-to-Book Ratio, Leverage, and Total Assets, are closely aligned across the two groups. This alignment is consistent with successful propensity score matching. Treated peers exhibit modestly higher Product Market Fluidity and Product Market Scope at the baseline, which is structurally expected given that they operate in the exact product space as a focal technology platform. These baseline differences in product space dynamics motivate the inclusion of Cohort-Time and Cohort-Firm fixed effects in the stacked difference-in-differences specification. Finally, consistent with the earlier validation in Table 1, the distributional statistics for the analytic weight $|\kappa_r|$ confirm a well-behaved, right-skewed,

[6] As a robustness check detailed in Appendix Table IA.3, I re-estimate the main specifications using three alternative empirical constructions, specifically the full unmatched sample, Coarsened Exact Matching, and Entropy Balancing, to verify that the core results are preserved. R&D expenditure remains fully robust across all four specifications, Sales is robust across three of the four, and ROA emerges as the most identification-sensitive outcome.

and approximately log-normal shape, which strictly supports its deployment as a continuous scalar weight in the main empirical specifications.

### 4.6 Empirical Strategy: Stacked Difference-in-Differences

To identify the causal impact of open-source software releases by focal technology firms, I employ a stacked difference-in-differences approach. This methodology directly addresses the potential biases inherent in standard two-way fixed effects models when treatment timing is staggered and treatment effects are heterogeneous across cohorts (Baker et al., 2022). The final dataset is constructed by concatenating the individual event cohorts into a single master stacked panel. To ensure comparability across heterogeneous firms and to mitigate the influence of extreme outliers, all continuous financial variables are normalized by total assets or total sales and winsorized at the 1st and 99th percentiles.

The primary empirical specification is modeled as follows:

$$Y_{i,c,t} = \alpha + \beta(Treat_{i,c} \times Post_{c,t}) + \Gamma X_{i,t} + \mu_{i,c} + \lambda_{c,t} + \varepsilon_{i,c,t} \tag{1}$$

where $Y_{i,c,t}$ represents the outcome of interest for firm $i$ in cohort $c$ at time $t$. A cohort $c$ is defined by a specific eligible open-source release event, spanning an observation window from $t = -3$ to $t = +3$ years relative to the release year. The interaction term $Treat_{i,c} \times Post_{c,t}$ is the primary variable of interest. $Treat_{i,c}$ is an indicator equal to one if firm $i$ operates in the same TNIC product market as the focal releasing firm in the pre-treatment year, representing the exposed competitor. $Post_{c,t}$ is an indicator taking the value of one for the periods following the open-source release where $t \geq 0$. To account for the heterogeneous economic significance of each release, the observations in this specification are weighted by the absolute value of the cohort-level Kogan intensity, $|\kappa_c|$, as defined in Section 4.1.

The vector $X_{i,t}$ represents a set of time-varying firm-level controls, including Firm Size, the Market-to-Book ratio, Return on Assets, and Leverage, ensuring that the estimates are not driven by shifting financial conditions. I control for unobservable confounders by including cohort-firm fixed effects, $\mu_{i,c}$, which absorb time-invariant firm characteristics within each specific event stack. Furthermore, I include cohort-time fixed effects, $\lambda_{c,t}$, to strictly control for macroeconomic shocks and time-series trends specific to each event window. Standard errors are clustered at the firm level to account for within-firm serial correlation over time.

# 5 Empirical Results

This section tests the mechanism chain laid out in Hypotheses 3 through 6 of Section 3. Because the specifications in this section pool substitute and complement peers, the estimates recover the average treatment effect over the full peer set. Under Proposition 2, this average is driven by the complement-peer subsample. The first subsection documents the top-line revenue compression predicted by Hypothesis 3. The second subsection shows that complement peers absorb the revenue shock by cutting direct costs, producing a short-run expansion in operating margins that completes Hypothesis 3. The third subsection traces the engineering labor reallocation toward the focal firm's open-source repositories predicted by Hypothesis 5. The fourth subsection documents the retreat in exploratory R&D and the aging of the patent portfolio predicted by Hypothesis 4. The final subsection shows that product-market fluidity contracts, completing Hypothesis 6. Section 6 then partitions the peer set along the Proposition 2 cut and confirms that the average effects documented below load on the complement-peer partition, with the substitute-peer subsample essentially flat.

## 5.1 The Revenue Shock: Impact on Peer Firm Sales

Table 3 summarizes how peer firm sales respond to the competitive shock of open-source entry. To ensure the robustness of this revenue impact, I present the results progressively from a baseline OLS model to a saturated specification that addresses both unobserved heterogeneity and the economic heterogeneity of the event. In column (1), using a baseline difference-in-differences specification without fixed effects, I find that normalized sales decrease by 0.049 in the treated group during the post-period compared to the control group, statistically significant at the 1% level.

To rigorously absorb unobservable confounders, I introduce strict Cohort-Firm and Cohort-Time fixed effects and firm-level controls in column (2), with standard errors clustered at the firm level. Under this saturated model, sales decrease by 0.046 in the treated group relative to the control group, statistically significant at the 5% level. In column (3), I further weight each observation by the absolute value of the cohort-level Kogan (2017) intensity, $|\kappa_c|$, so that more economically significant releases receive greater weight in the estimation. The Kogan-weighted coefficient is $-0.060$, statistically significant at the 1% level. Economically, a 0.060 reduction in sales scaled by total assets is substantial. Relative to the sample mean of 0.728, this translates to an 8.2% decline in the overall revenue generation capacity of incumbent competitors, with the magnitude increasing monotonically as the specification places more weight on the most consequential releases. This top-line compression confirms that when dominant platforms release high-quality open-source software, it acts as a direct substitute that cannibalizes the market share and pricing power of proprietary alternatives.

[Insert Table 3 Here]

To validate the causal interpretation of this revenue shock, Figure 3 plots the dynamic treatment effect on sales relative to the base year. The visual evidence supports the parallel trends assumption essential for the difference-in-differences framework. In the pre-treatment years, the estimated coefficients are statistically indistinguishable from zero, indicating no divergent pre-existing trends between the treated and control groups prior to the event. Following the open-source release at year zero, there is a sharp and persistent downward trajectory in peer sales. This dynamic plot confirms that the revenue compression is not driven by prior firm distress but is a direct consequence of the focal firm's open-source entry.

[Insert Figure 3 Here]

## 5.2 Defensive Restructuring: Cost Efficiency and Profitability

An 8.2 percent structural decline in sales presents a substantial operational challenge to incumbent firms. The economics of open-source software provides a mechanism for these firms to absorb this top-line shock. Theoretical and empirical literature suggests that open-source releases generate substantial technological spillovers, allowing rival firms to free-ride on the focal platform's R&D investments (Nagle, 2019; Emery et al., 2025). By adopting the newly released open-source infrastructure, incumbent firms can replace expensive proprietary software inputs, eliminate redundant development efforts, and achieve significant technological cost efficiencies (Wen et al., 2016; Haese and Peukert, 2025).

To understand how peer firms leverage this opportunity to defend their bottom line, Table 4 examines the shifts in their underlying cost structures and resultant profitability metrics. The table presents two outcomes, Return on Assets (ROA) and Cost of Goods Sold (COGS), each estimated under a baseline OLS specification and the saturated specification with Cohort-Firm and Cohort-Time fixed effects, firm-level controls, and Kogan-intensity weights.

I first analyze the profitability result in columns (1) and (2). The baseline OLS coefficient in column (1) is 0.018, statistically significant at the 1% level. In column (2), the saturated specification yields a coefficient of 0.053, statistically significant at the 5% level. This implies a 5.3 percentage point improvement in overall asset profitability for treated peer firms in the post-period. The sign and magnitude of this effect establish that peer firms absorb the revenue shock without allowing their bottom line to deteriorate; in fact, short-term profitability expands.

Columns (3) and (4) turn to the direct cost channel that mechanically drives this margin expansion. In the baseline column (3), COGS scaled by total assets decreases by 0.045, statistically significant at the 1% level. In the saturated column (4), the coefficient strengthens to -0.054, also significant at the 1% level. The 0.054 estimate translates to a 5.4 percentage point reduction in normalized direct costs, representing a 16.5% decline relative to the sample mean of 0.328. This reduction confirms that peer firms systematically scale back their variable expenses, capitalizing on the technological spillovers from the free infrastructure to optimize cost efficiency. The profitability expansion in columns (1)–(2) is therefore not driven by top-line recovery but by this aggressive compression of the cost base.

[Insert Table 4 Here]

To validate the sequence of this defensive restructuring, Figure 4 plots the dynamic treatment effects for both ROA and COGS. The evidence supports the parallel trends assumption, as the pre-treatment coefficients for both variables remain flat and statistically indistinguishable from zero prior to the event year. Following the open-source release, Panel B illustrates a persistent downward trajectory in the cost of goods sold over the subsequent three years. Mirroring this timeline, Panel A displays a consistent upward trajectory in ROA. This dynamic symmetry provides evidence that the observed margin expansion is not driven by organic business growth but is the direct consequence of shrinking the firm's cost base via open-source adoption in response to the exogenous revenue shock.

[Insert Figure 4 Here]

## 5.3 Labor Co-optation: Developer Mobilization Toward the Focal Platform

The defensive cost-cutting documented in the previous subsection raises an immediate structural question. If peer firms are preserving their product lines while simultaneously slashing internal overhead, where does the engineering workforce go? Labor reallocation theories suggest that dominant platforms can co-opt the proprietary talent of their rivals by setting open technological standards (Gupta et al., 2024). To maintain product compatibility with the new industry baseline, rivals are economically incentivized to redirect their developers to the platform's open-source ecosystem (Lerner and Tirole, 2002). This subsection directly tests that mechanism using real-time developer engagement on GitHub.

Table 5 empirically tests this mechanism across two complementary samples. Panel A reports the full sample of approximately 1,700 Kogan-eligible cohorts, the broader peer universe in which the average response is estimated. Panel B restricts the analysis to the Top 100 BigQuery-linked sub-sample, in which contributor identity is most precisely linked to firm-level employment; this is the identification sample matching the dynamic event-study in Figure 5. Both panels weight observations by the absolute Kogan (2017) innovation value so that the estimates reflect the response to the most economically influential releases. Cohort $\times$ year fixed effects absorb the secular GitHub-adoption trend that affects all peer firms during 2014–2017.[7]

In Panel A, total developer engagement, measured by any work per developer, increases by 1.145 in the treated group during the post-period relative to the control group, statistically significant at the 1% level (column 1). This represents a 36.7% increase relative to the full-sample mean of 3.122 contributions per developer. The mobilization is driven almost entirely by active structural contributions: in column (2), active contributions (pushes, pull requests, issues, issue comments, and forks of the focal repository) increase by 1.134, statistically significant at the 5% level, accounting for virtually the entire change in total engagement. The active developer base also

[7] Unweighted variants of both panels, in which each cohort enters the estimate equally, are reported in Internet Appendix Table IA.7; the unweighted specifications are larger in magnitude and tighter in significance, indicating that the average response across all releases is at least as strong as the response to the most influential releases.

expands meaningfully: in column (4), the log number of working developers increases by 0.141 in the treated group, statistically significant at the 1% level, corresponding to a 15.1% increase in active development staffing dedicated to the focal open-source repositories. The passive-engagement coefficient in column (3) is statistically indistinguishable from zero in the Kogan-weighted specification, indicating that the response is concentrated in active rather than passive engagement. Panel B confirms the same qualitative pattern within the Top 100 BigQuery-linked sub-sample: any work per developer increases by 0.347 (column 1, $p < 0.05$) and active contributions by 0.341 (column 2, $p < 0.05$), reaffirming that the headline result is not an artifact of the broader peer universe. The relative magnitudes across the two panels make clear that the labor co-optation effect operates both at scale (Panel A) and within the BigQuery-linked identification sample where contributor identity is most precisely measured (Panel B).

[Insert Table 5 Here]

To validate the timing of this labor reallocation, Figure 5 plots the dynamic treatment effects on per-developer contributions for the Top 100 identification sample. The visual evidence supports the parallel-trends assumption: the pre-treatment coefficients in both Panel A (any work per developer) and Panel B (active contributions per developer) are statistically indistinguishable from zero. Immediately following the open-source release, both panels exhibit a sharp upward trajectory in developer engagement that peaks at the event year. This confirms that the mobilization of engineering talent is a direct response to the exogenous technological shock rather than a pre-existing trend among peer firms.

[Insert Figure 5 Here]

## 5.4 The Long-Term Cost: Stagnation in Exploratory Innovation

The labor reallocation documented above does not merely relocate developer hours; it structurally hollows out the rival's capacity for independent discovery. As engineering effort migrates from proprietary projects to the focal platform's open-source commons, the firm's internal pipeline for frontier research contracts. When incumbent firms substitute proprietary development with external open-source public goods, the reliance on external knowledge often crowds out internal research and development (Cassiman and Veugelers, 2006). Consequently, this divestment leads to a decay in the rival's independent technological trajectory, forcing a structural shift away from exploratory discovery and toward exploitative innovation (Benner and Tushman, 2002). Table 6 tests this theoretical hypothesis by examining the shifts in peer firms' innovation activities following the competitive shock.

Table 6 reports three innovation outcomes, R&D expenditure scaled by total assets, the average age of the firm's overall patent portfolio, and the average age of its exploratory patent portfolio. Each outcome is estimated under a baseline OLS specification and a saturated specification with Cohort-Firm and Cohort-Time fixed effects, firm-level controls, and Kogan-intensity weights.

In columns (1) and (2), I find that R&D expenditures scaled by total assets decrease in the treated group during the post-period compared to the control group. The baseline OLS coefficient in column (1) is $-0.006$, significant at the 1% level. The saturated specification in column (2) produces a larger estimate of $-0.014$, significant at the 1% level. This translates to a 1.4 percentage point drop in R&D intensity, or a 13.1% decline in research investment relative to the sample mean of 0.107. This reduction confirms that treated firms actively scale back their proprietary technological investments once a free alternative becomes available.

With less capital allocated to internal research, the firm's innovation portfolio ages. In columns (3) and (4), I find that the average age of the overall patent portfolio increases in the treated group during the post-period compared to the control group, with a saturated-specification coefficient of 0.323 in column (4), significant at the 1% level. This aging process is particularly evident at the technological frontier. In columns (5) and (6), the average age of exploratory patents increases by 0.792 years in column (6) in the treated group relative to the control group, also significant at the 1% level. By commoditizing the leading edge of technology, focal platforms force their rivals to slow the introduction of new patents, allowing both the overall and the exploratory portfolio to age measurably within the three-year post-event window.

[Insert Table 6 Here]

To validate the causal nature of this innovation decay, Figure 6 plots the dynamic treatment effects for R&D expenditures and average patent portfolio age. The pre-treatment coefficients for both variables remain flat and statistically indistinguishable from zero prior to the event. Following the open-source release, Panel A reveals a persistent downward trajectory in R&D investment over the subsequent three years. Mirroring this structural decline, Panel B displays a continuous upward trend in the average age of the patent portfolio. This dynamic evidence indicates that the effect on innovation is a direct consequence of the focal firm's open-source entry.

[Insert Figure 6 Here]

## 5.5 Freezing the Competitive Frontier: Product Market Dynamics

If incumbent rivals are actively reducing their internal proprietary R&D, this internal structural shift must manifest in their actual product offerings. Theoretical models of product differentiation suggest that when firms rely on standardized external infrastructure, their product evolution stagnates. To capture this external market positioning, Table 7 utilizes the text-based network industry classifications developed by Hoberg et al. (2014). Specifically, I examine three textual measures extracted from the product descriptions in annual 10-K filings: Product Market Fluidity, Self-Fluidity, and Product Market Scope. Product Market Fluidity gauges how rapidly a firm's product vocabulary changes relative to the aggregate shifts in its competitors' vocabularies. Self-Fluidity isolates the firm's own year-over-year vocabulary changes. Product Market Scope measures the overall breadth of the firm's product lines. By analyzing these continuous metrics, this section tests whether the open-source shock freezes the competitive frontier.

In column (1), I find that Product Market Fluidity decreases in the treated group during the post-period compared to the control group, statistically significant at the 1% level. The coefficient of -0.326 represents a 6.2% decline in market fluidity relative to its sample mean of 5.296. Because this metric measures changes relative to competitors, the reduction indicates that treated firms are falling behind dynamic shifts in their industry.

This stagnation is driven by a halt in their internal development efforts. In column (2), I find that Self-Fluidity decreases in the treated group during the post-period compared to the control group, statistically significant at the 1% level. The coefficient of -2.271 translates to an 11.1% decline in the firm's own product differentiation efforts relative to its mean of 20.41. In column (3), the coefficient for Product Market Scope is statistically indistinguishable from zero. This combination of results suggests a chilling effect where incumbent firms do not necessarily shrink their product lines, but they freeze the evolution and differentiation of their existing products (Bloom et al., 2013).

[Insert Table 7 Here]

To validate the causal nature of this product market effect, Figure 7 plots the dynamic treatment effects for both Product Market Fluidity and Self-Fluidity. The pre-treatment coefficients in both panels remain relatively flat and statistically indistinguishable from zero. Immediately following the open-source release, both Panel A and Panel B exhibit a persistent downward trajectory. This dynamic visualization confirms that the contraction in product differentiation is a direct consequence of the focal platform's open-source entry.

[Insert Figure 7 Here]

# 6 Substitute vs. Complement Peers

The baseline results establish a robust average treatment effect. Focal firms' open-source releases trigger revenue compression, defensive cost-cutting, labor co-optation, innovation decay, and a freezing of product market dynamics among incumbent peers. This section tests the central theoretical claim that this competitive displacement is targeted specifically at the complementary layer of the platform rather than at direct substitutes in the focal firm's core market. I partition the treated peer cohort into Substitute and Complement groups, compare the magnitude of the real-side shocks across the two segments, and then show that the stock market immediately prices the release as a wealth transfer flowing almost entirely out of the complement pool. Together, these analyses complete the paper's causal chain from product-market exposure to investor-implied valuation.

## 6.1 Classifying Peers into Substitutes and Complements

The distinction between a substitute and a complement is constructed from two metrics that characterize the economic relationship between each peer firm and the releasing repository. The first metric is technology overlap. Following the foundational approach of Jaffe (1986) and subsequent literature on technological proximity (Bloom et al., 2013; Bena and Li, 2014), this overlap is measured as the TF-IDF cosine similarity between the Cooperative Patent Classification (CPC) profile of the peer firm's patent portfolio and the CPC profile inferred for the released open-source repository. I restrict the heterogeneity sample to peers in the top half of the CPC-overlap distribution, isolating firms whose products are technologically exposed to the release.

The second metric is the sign of the peer's stock-return co-movement with the focal firm, measured over the pre-event window $[t-250, t-30]$ after netting out industry returns. The use of return co-movement to identify economically linked firm pairs follows a well-established literature on information transmission and production complementarity (Lee et al., 2024). A peer whose returns move in the same direction as the focal firm is classified as a Complement. In this case, the market prices the release as joint news for both firms, consistent with the peer supplying a complementary layer to the focal platform. A peer whose returns move in the opposite direction is classified as a Substitute. The market treats the release as a zero-sum reallocation for these firms, with gains to the focal firm offsetting losses at the peer.

This classification directly tests the theoretical prediction. While traditional competitive models suggest the strongest revenue impact should fall on direct substitutes, the commoditization of the complement theory predicts the opposite. The open-source release is designed to erode the pricing power of firms whose products were complementary inputs to the focal platform, converting paid complements into free alternatives.

## 6.2 Differential Impact on Sales and Direct Costs

Table 8 reports the decomposition for Sales and COGS. The specification is identical to the pooled baseline, estimated separately within each subsample. Columns (1) and (2) report the Sales results. For Substitute peers, the treatment coefficient is -0.034, which is statistically indistinguishable from zero. For Complement peers, the coefficient is -0.093, statistically significant at the 1% level. The magnitude of the decline is nearly 2.8 times larger for Complement peers. This direction confirms that the peers whose proprietary offerings moved with the focal firm prior to the release experience the most severe revenue compression.

Columns (3) and (4) report the decomposition for COGS. Both groups cut direct costs post-release, but the Complement peer cut of -0.064 is substantially larger than the Substitute peer cut of -0.038. This asymmetry is economically intuitive. The free infrastructure released by the focal firm is a generalized input that any technologically adjacent peer can substitute into its cost base. However, only Complement peers lose the paid pricing surface that the release directly commoditizes.

[Insert Table 8 Here]

The dynamic counterpart of these estimates is shown in Figure 8. Panel A plots the event-study trajectory for Complement peers, and Panel B plots it for Substitutes. The Complement peer sales path is flat prior to the release, drops sharply at year zero, and remains persistently below the pre-period base. The Substitute peer path shows no meaningful deviation from zero. This graphical evidence localizes the full weight of the sales effect documented in the baseline results to the Complement partition.

[Insert Figure 8 Here]

## 6.3 Market Reaction and Wealth Transfer

The real-side results establish that the revenue and cost restructuring fall overwhelmingly on complement peers. If investors correctly understand the economics of commoditizing the complement, they should reprice the affected firms along the same partition in real time. This subsection tests that prediction by examining the cumulative abnormal returns (CAR) of focal firms, substitute peers, and complement peers in a narrow window around the release, and by translating these percentage returns into aggregate dollar-value impacts. Together, these estimates complete the causal chain by showing that the market's immediate pricing of the shock mirrors the product-market partition documented above.

Before presenting the estimates, a sample restriction specific to the high-frequency market-reaction analysis is necessary. The daily FF4 factor loadings required for firm-specific abnormal returns are estimated on a 150-day window ending 30 days before each release. Extending this estimation across all approximately 1,700 releases for every product-market peer is not feasible with the current CRSP-linked return pipeline. I therefore focus the market-reaction analysis on the top 100 strategically influential releases, which collectively cover the overwhelming majority of the Kogan-weighted intensity mass in the full sample. Pre-trends remain statistically flat in all three partitions, confirming that this restriction does not induce look-ahead bias. The full-sample robustness check reported in Appendix Section B.2 reproduces the same qualitative pattern using every product-market peer for which return data are available, with the aggregate dollar-value transfer to the complement pool surviving intact.

To reflect the economic weight of each firm rather than treating a \$1 billion peer as equivalent to a \$1 trillion peer, I compute value-weighted CARs across the three groups. Firms are weighted by their pre-event market capitalization, and standard errors are constructed from the effective-sample-size weighted cross-sectional variance.[8] Table 9 reports the value-weighted three-day CARs over the $[-1, +1]$ window centered on the release. Focal firms gain 0.46 percentage points, significant at the 10% level. Substitute peers exhibit a point estimate of $+0.08$ percentage points that is statistically indistinguishable from zero. Complement peers lose 2.81 percentage points, significant at the 1% level. Economically, this single-day estimate implies that the market reassesses the present value of complement firms' expected future cash flows at a magnitude roughly six times the focal premium, confirming the targeted nature of the valuation shock.

[Insert Table 9 Here]

Figure 9 visualizes this three-way partition as value-weighted bars with 95% confidence intervals. The graphical evidence immediately communicates the asymmetry of the $[-1, +1]$ reaction. The focal gain and the complement loss are both statistically separated from zero and from each other, while the substitute bar collapses onto zero. Economically, the figure compresses the central message of the market-reaction analysis into a single image: the release reprices complements strongly negatively, reprices focal firms modestly positively, and leaves substitutes unaffected, exactly as the commoditize-your-complement hypothesis predicts.

[Insert Figure 9 Here]

[8]The analogous equal-weighted estimates, the corresponding $t$-statistics, and a direct comparison between the two weighting schemes are reported in Appendix Section B.2. The value-weighted framework is the economically correct one here because the outcome of interest, aggregate wealth destruction in the complement pool, is intrinsically dollar-weighted, whereas the equal-weighted version conflates the large-cap incumbents that bear most of the economic impact with a long tail of small peers whose dollar stakes are negligible.

To verify that this three-day point estimate is not an artifact of a noisy reaction window, Figure 10 extends the analysis to the full daily trajectory. The figure traces the value-weighted cumulative average abnormal return of focal firms and complement peers from day $-10$ through day $+20$, with 95% confidence bands constructed from the weighted cross-sectional variance. The placebo window $[-10, -2]$ is flat for both groups, with $t$-values of $+0.90$ and $-0.16$, respectively, providing the clean pre-trend required to attribute the post-event separation to the release itself. From day $-1$ onward, the two trajectories diverge sharply and remain separated throughout the 20-day post-window, with the complement path stabilizing at roughly $-3.1$ percentage points and the focal path at approximately $+2.7$ percentage points by day $+20$. The persistence of the separation rules out mechanical post-event bid-ask bounce or short-horizon microstructure noise as an explanation for the headline estimate.

[Insert Figure 10 Here]

To translate these percentage returns into economic magnitudes, I follow recent work evaluating the financial footprint of open-source innovation (Emery et al., 2025; Hoffmann et al., 2024) and multiply each firm's estimated CAR by its market capitalization on the trading day immediately preceding the release. Figure 11 aggregates this dollar impact across the three partitions. In the $[-1, +1]$ window, focal firms gain an aggregate \$313 billion, substitute peers gain a statistically insignificant \$84 billion, and complement peers lose \$153 billion. The focal gain and the complement loss together account for virtually the entire gross wealth movement, confirming that investors price these releases as a directed transfer from the complement pool to the releasing platform rather than as a diffuse industry spillover.

[Insert Figure 11 Here]

Taken together, the market-reaction evidence completes the mechanism laid out in the preceding subsections. The same complement firms that suffer the largest sales compression and execute the sharpest cost restructuring are the firms whose equity value the market marks down in real time. Substitute peers, whose real-side outcomes are economically muted, receive an economically negligible market reaction. This alignment between the product-market partition and the contemporaneous market-implied valuation shock constitutes the paper's strongest evidence that open-source releases operate as targeted commoditization events rather than as benign technological spillovers.

## 6.4 Discussion and Robustness

To provide concrete economic context for this partition, Appendix Table IA.5 presents illustrative examples of firm pairs categorized into the Substitute and Complement buckets. A qualitative review of these matched pairs confirms that the algorithmic classification accurately aligns with the structural realities of the product market. Firms classified as Complements typically provide adjacent or interdependent services to the focal platform, whereas those classified as Substitutes offer direct, competing alternatives.

A potential econometric concern with the baseline partition, however, is that the sign of pre-event return co-movement might serve as a noisy proxy for the underlying economic relationship. To address this concern, Appendix Table IA.6 replaces the one-dimensional sign split with a rigorous two-dimensional classification. This strict definition requires Pure Complement peers to satisfy both high CPC overlap and positive return co-movement, while Pure Substitutes must satisfy high CPC overlap and negative return co-movement. The Sales and COGS results remain unchanged in direction and statistical significance under this stringent classification, confirming that the concentration of the revenue shock among complement firms is robust to alternative definitions.

This decomposition refines the paper's central empirical finding. The average treatment effects documented in the baseline analysis do not represent a diffuse industry spillover. Instead, they represent a targeted compression of peer firms whose products operate in a complementary relationship to the releasing platform. Furthermore, the empirical results rule out conventional head-to-head rivalry as the primary channel of cannibalization. Substitute peers absorb a substantially smaller shock than Complement peers do. Ultimately, these focal platforms successfully commoditize the paid complements upon which their own ecosystems previously relied.

A final concern with the baseline DiD specification is that the two-way fixed-effect estimator can be biased under staggered treatment timing and heterogeneous treatment effects, because already-treated units serve as implicit controls for later-treated cohorts (Goodman-Bacon, 2021; Sun and Abraham, 2021). To address this, I re-estimate the focal-firm DiD using the Callaway and Sant'Anna (2021) doubly-robust ATT estimator with a never-treated control group. This estimator weights only clean 2×2 cohort-time comparisons and excludes the forbidden comparisons that drive the TWFE bias. For this purpose, I aggregate the approximately 1,700 stacked release events to firm-level treatment timing by assigning each focal firm the year of its first eligible open-source release, yielding eight treated firms (Adobe, Alphabet, Apple, Meta, Microsoft, Netflix, and Nvidia treated in 2015, and Oracle treated in 2018) and 443 never-treated control firms over fiscal years 2010 through 2024. Table 11 reports the results for the two headline financial outcomes. For Return on Assets, the TWFE coefficient is 0.052 ($t = 2.20$); the Callaway–Sant'Anna ATT is 0.131 ($t = 4.95$). For Log Sales, the TWFE coefficient is 0.431 ($t = 1.89$); the Callaway–Sant'Anna ATT is 1.040 ($t = 3.94$). The sign is preserved on both outcomes, but the CS estimates are roughly twice the size of the TWFE estimates and are tighter, indicating that the paper's main TWFE estimates are a conservative lower bound on the true treatment effect.

[Insert Table 11 Here]

Figure 13 reports the dynamic event-study counterpart. The pre-treatment coefficients lie within the confidence band around zero for both outcomes, supporting the parallel-trends assumption; the post-treatment trajectory is positive and monotonically increasing through year nine, with ROA significant from event time $+1$ onward.

[Insert Figure 13 Here]

# 7 Conclusion

The modern digital economy is increasingly characterized by a structural shift wherein dominant technology firms invest substantial capital to develop advanced software, only to release it to the public at no cost. While the open-source movement is frequently viewed as a collaborative public good, this study provides comprehensive empirical evidence of a more calculated economic reality. By integrating the economics of two-sided network effects with the resource-based view of the firm, this paper demonstrates that dominant platforms strategically deploy open-source software to commoditize complementary markets, displacing the proprietary resource advantages of incumbent firms.

The empirical anatomy of this commoditization shock is sequential and structurally significant. Confronted with a structural top-line revenue compression, incumbent peers are forced into an operational restructuring. They reduce direct costs and organizational overhead by substituting proprietary inputs with the newly available open-source infrastructure. While this defensive pivot mechanically inflates short-term profit margins, it is not a costless adjustment. Instead, it is mirrored in a substantial reallocation of engineering human capital, as competitor engineers are effectively co-opted and mobilize on platforms such as GitHub to maintain and expand the focal firm's open-source ecosystem.

By outsourcing their core technological foundations to the dominant platform in this way, these rivals systematically hollow out their independent innovation capabilities. Research and development budgets contract, patent portfolios age rapidly, and exploratory innovation at the technological frontier declines. The rival's proprietary product lines stagnate, freezing their product market differentiation.

The heterogeneity analysis confirms that these dynamics are not the byproduct of a diffuse technological spillover. The impact of open-source releases is highly targeted. When partitioning the treated cohort into substitute and complement peers, the revenue compression, cost restructuring, and labor mobilization are heavily concentrated among firms operating as complements to the focal platform, while direct substitutes remain largely insulated. This targeted partition is immediately validated by the financial markets: investors price each release as a real-time wealth transfer from the complement pool to the releasing platform, with the focal firm gaining \$313 billion in aggregate market value and the complement pool losing \$153 billion within the three-day event window. These asset-price estimates close the causal chain by showing that the product-market partition and the contemporaneous market-implied valuation shock line up on the same firms, confirming the central theoretical prediction that open-source software is strategically targeted to erode the pricing power of complementary layers rather than head-to-head substitutes.

The historical precedent for this dynamic clarifies what is at stake. Gurley (2026) notes that Amazon Web Services captured the cloud only because open-source commoditization of the operating-system and middleware stack had already eliminated the supplier power that would have allowed IBM, Hewlett-Packard, or Intel to extract rents. In that case the commoditization was exogenous, driven by a developer community rather than by Amazon. The contemporary version of the same mechanism, documented here, is endogenous: dominant incumbents deliberately release the complementary asset themselves, financed out of profit-center rents, to occupy the operator position. The redistributive consequences are the same, but the strategic agency has changed, and the regulatory framework has not caught up.

These findings carry significant implications for both academic theory and antitrust policy. In the strategic management literature, this paper illustrates how cross-market subsidies can be utilized to convert a rival's proprietary assets into obsolete legacy systems. Furthermore, it highlights a critical blind spot in modern antitrust enforcement. Traditional regulatory frameworks focus predominantly on consumer harm via monopoly pricing and output restrictions. However, in the digital platform ecosystem, market power is frequently exercised through the free distribution of complements. By giving away advanced software, dominant platforms avoid traditional antitrust scrutiny while successfully diminishing rivals' revenue, stifling independent exploratory innovation, and absorbing the industry's engineering human capital. Ultimately, this study demonstrates that through the strategic commoditization of the complement, dominant platforms systematically neutralize competitive threats and secure their structural advantage in the digital economy.

# References


Adner, Ron, and Rahul Kapoor, 2010, Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations, *Strategic Management Journal* 31, 306–333.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, 2005, Competition and innovation: An inverted-u relationship, *The Quarterly Journal of Economics* 120, 701–728.

Alexy, Oliver, Gerard George, and Ammon J. Salter, 2013, Cui bono? the selective revealing of knowledge and its implications for innovative activity, *Academy of Management Review* 38, 270–291.

Alexy, Oliver, Joel West, Helge Klapper, and Markus Reitzig, 2018, Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm, *Strategic Management Journal* 39, 1704–1727.

Armstrong, Mark, 2006, Competition in two-sided markets, *RAND Journal of Economics* 37, 668–691.

Baker, Andrew C, David F Larcker, and Charles CY Wang, 2022, How much should we trust staggered difference-in-differences estimates?, *Journal of Financial Economics* 144, 370–395.

Baker, Scott R, Brian Baugh, and Marco Sammon, 2023, Customer churn and intangible capital, *Journal of Political Economy Macroeconomics* 1, 447–505.

Balsmeier, Benjamin, Lee Fleming, and Gustavo Manso, 2017, Independent boards and innovation, *Journal of Financial Economics* 123, 536–557.

Barney, Jay, 1991, Firm resources and sustained competitive advantage, *Journal of Management* 17, 99–120.

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, *The Journal of Finance* 69, 1923–1960.

Benner, Mary J, and Michael Tushman, 2002, Exploratory and exploitative innovation, learning, and the impact of corporate quality management, *Academy of Management Journal* 45, 676–698.

Bian, B., Q. Huang, Y. Li, and H. Tang, 2025, Data as a networked asset, *Working Paper* .

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, 2013, Identifying technology spillovers and product market rivalry, *Econometrica* 81, 1347–1393.

Boudreau, Kevin, 2010, Open platform strategies and innovation: Granting access vs. devolving control, *Management Science* 56, 1849–1872.

Callaway, Brantly, and Pedro HC Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225, 200–230.

Casadesus-Masanell, Ramon, and Pankaj Ghemawat, 2006, Dynamic mixed duopoly: A model motivated by Linux vs. Windows, *Management Science* 52, 1072–1084.

Casadesus-Masanell, Ramon, and Gaston Llanes, 2011, Mixed source, *Management Science* 57, 1212–1230.

Cassiman, Bruno, and Reinhilde Veugelers, 2006, In-house r&d and external knowledge acquisition: complementary or substitute?, *Research Policy* 35, 68–82.

Cennamo, Carmelo, Hakan Ozalp, and Tobias Kretschmer, 2018, Platform architecture and quality trade-offs of multihoming complements, *Information Systems Research* 29, 461–478.

Chesbrough, Henry, and A. K. Crowther, 2006, Beyond high tech: early adopters of open innovation in other industries, *R&D Management* 36, 229–236.

Coleman, J., K. Fronk, and K. Valentine, 2025, Corporate adoption of an open innovation strategy: Evidence from github, *Working Paper* .

Conti, Annamaria, Christian Peukert, and Maria Roche, 2025, Beefing it up for your investor? engagement with open source communities, innovation, and startup funding: Evidence from github, *Organization Science* 36, 1551–1573.

Economides, Nicholas, and Evangelos Katsamakas, 2006, Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry, *Management Science* 52, 1057–1071.

Eisenmann, Thomas, Geoffrey Parker, and Marshall Van Alstyne, 2011, Platform envelopment, *Strategic Management Journal* 32, 1270–1285.

Emery, Logan P., Chan Lim, and Shiwei Ye, 2025, The private value of open-source innovation, Working paper, SSRN.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Farrell, Joseph, and Garth Saloner, 1985, Standardization, compatibility, and innovation, *RAND Journal of Economics* 16, 70–83.

Fosfuri, Andrea, Marco S. Giarratana, and Alessandra Luzzi, 2008, The penguin has entered the building: The commercialization of open source software products, *Organization Science* 19, 292–305.

Frank, Murray Z, and Vidhan K Goyal, 2009, Capital structure decisions: which factors are reliably important?, *Financial Management* 38, 1–37.

Gawer, Annabelle, and Rebecca Henderson, 2007, Platform owner entry and innovation in complementary markets: Evidence from intel, *Journal of Economics & Management Strategy* 16, 1–34.

Goodman-Bacon, Andrew, 2021, Difference-in-differences with variation in treatment timing, *Journal of Econometrics* 225, 254–277.

Gupta, A., N. Nishesh, and E. Simintzi, 2024, Big data and bigger firms: A labor market channel, *SSRN Electronic Journal* .

Gurley, Bill, 2026, From Open Source Software to Open Source Strategy: How the Smartest Executives Are Using Open Source Techniques to Optimize Corporate Strategy, `https://p3institute.substack.com/p/from-open-source-software-to-open`, Substack post, May 10, 2026.

Haese, M., and C. Peukert, 2025, Open at the core: Moving from proprietary technology to building a product on open source software, *Working Paper* .

Heath, D., N. Seegert, R. Wuebker, and J. Yang, 2023, Teams and the homophily trap: Evidence from open source software, *Working Paper* .

Henkel, Joachim, 2006, Selective revealing in open innovation processes: The case of embedded Linux, *Research Policy* 35, 953–969.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

Hoberg, Gerard, Gordon Phillips, and Nagpurnanand Prabhala, 2014, Product market fluidities, technology spillovers, and product market rivalry, *The Review of Financial Studies* 27, 333–371.

Hoffmann, Manuel, Frank Nagle, and Yanuo Zhou, 2024, The value of open source software, *SSRN Electronic Journal* .

Jaffe, Adam B, 1986, Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value, *The American Economic Review* 76, 984–1001.

Katz, Michael L., and Carl Shapiro, 1985, Network externalities, competition, and compatibility, *American Economic Review* 75, 424–440.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *The Quarterly Journal of Economics* 132, 665–712.

Lee, C. M. C., T. T. Shi, S. T. Sun, and R. Zhang, 2024, Production complementarity and information transmission across industries, *Journal of Financial Economics* 155, 103812.

Lerner, Josh, and Jean Tirole, 2002, Some simple economics of open source, *The Journal of Industrial Economics* 50, 197–234.

Lin, Y.-K., and L. M. Maruping, 2022, Open source collaboration in digital entrepreneurship, *Organization Science* 33, 212–230.

Makaew, Tanakorn, and Vojislav Maksimovic, 2020, Competition and operating volatilities around the world, *Journal of Financial and Quantitative Analysis* 55, 517–547.

Manso, Gustavo, 2011, Motivating innovation, *The Journal of Finance* 66, 2113–2153.

Mkrtchyan, A., J. Bai, R. Dai, and C. Wan, 2025, Creativity without walls: The case of open innovation, *SSRN Electronic Journal* .

Nagaraj, Abhishek, and Henning Piezunka, 2024, The divergent effects of firm sponsorship on open collaboration, *Strategy Science* 9, 277–296.

Nagle, Frank, 2019, Open source software and firm productivity, *Management Science* 65, 1191–1215.

Parker, Geoffrey G, and Marshall W Van Alstyne, 2005, Two-sided network effects: A theory of information product design, *Management Science* 51, 1494–1504.

Roberts, Michael R, and Toni M Whited, 2013, Endogeneity in empirical corporate finance, in *Handbook of the Economics of Finance*, volume 2, 493–572 (Elsevier).

Rochet, Jean-Charles, and Jean Tirole, 2003, Platform competition in two-sided markets, *Journal of the European Economic Association* 1, 990–1029.

Seamans, Robert, and Feng Zhu, 2014, Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers, *Management Science* 60, 476–493.

Shapiro, Carl, and Hal R. Varian, 1999, *Information Rules: A Strategic Guide to the Network Economy* (Harvard Business School Press, Boston, MA).

Spolsky, Joel, 2002, Strategy letter V, Essay, Joel on Software, `https://www.joelonsoftware.com/2002/06/12/strategy-letter-v/`.

Sun, Liyang, and Sarah Abraham, 2021, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Journal of Econometrics* 225, 175–199.

Wen, Wen, Marco Ceccagnoli, and Chris Forman, 2016, Opening up intellectual property strategy: Implications for open source software entry by start-up firms, *Management Science* 62, 2668–2691.

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

Zhao, Yue, 2025, Open-source generative ai and firm value: Evidence from the release of deepseek, *Working Paper, Singapore Management University* .

Zhu, Kevin Xiaoguo, and Zach Zhizhong Zhou, 2012, Research note—lock-in strategy in software competition: Open-source software vs. proprietary software, *Information Systems Research* 23, 536–545.

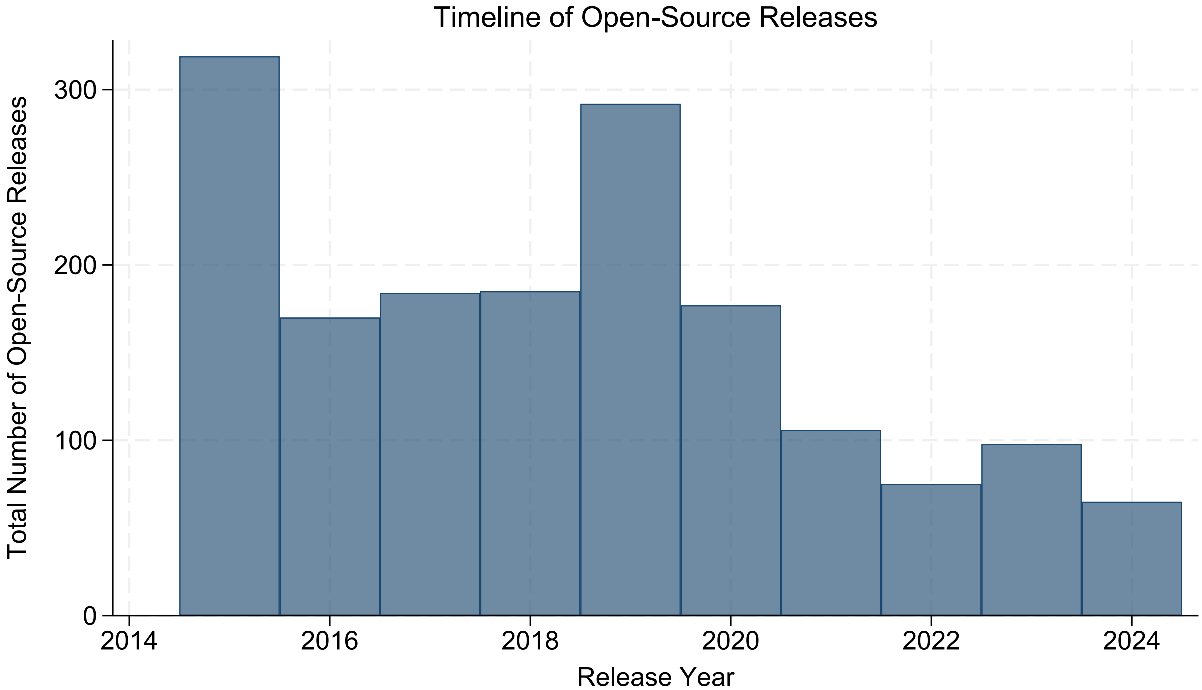
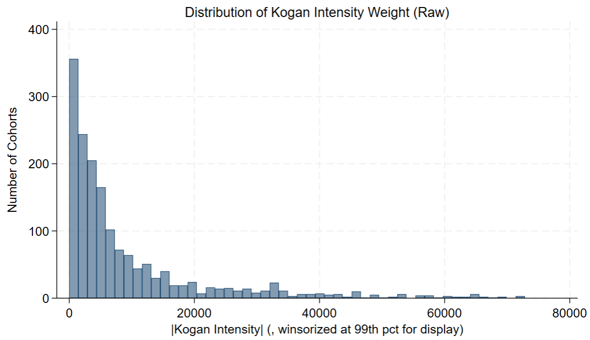
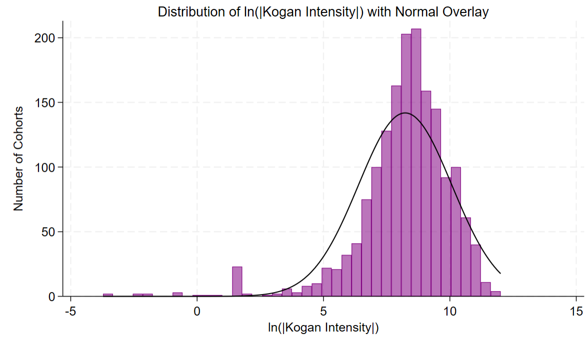



**Figure 1. Timeline of Total Open-Source Releases**
This figure presents the annual distribution of the approximately 1,700 open-source repositories in the full eligible cohort set released between 2015 and 2024. The sample utilizes the complete set of open-source releases, weighted by the cohort-level market-value reaction. The histogram displays the concentration of early releases from 2015 to 2017 alongside the distribution of qualifying releases across the entire sample period.




**(a)** Raw Absolute Innovation Value (winsorized at 99th pct)




**(b)** Log-Transformed Value with Normal Overlay

**Figure 2. Distribution of Cohort-Level Intensity Weights**
This figure plots the cohort-level distribution of the absolute Kogan (2017) innovation value used as the analytic weight in the stacked Difference-in-Differences (DiD) estimation. Panel (a) shows the raw distribution of the absolute intensity, winsorized at the 99th percentile for display clarity. Panel (b) displays the log-transformed distribution with a normal-density overlay.

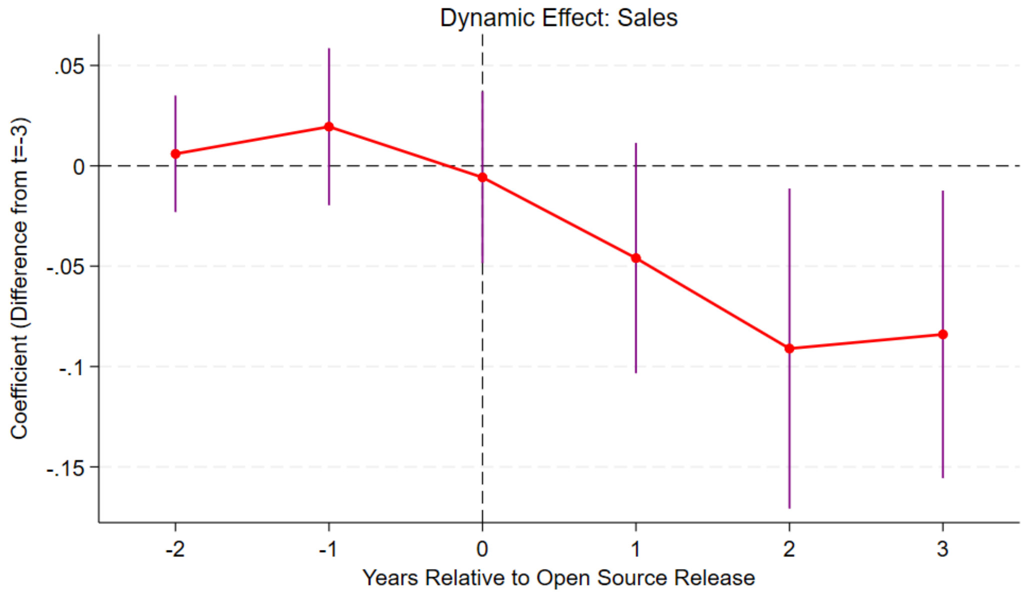
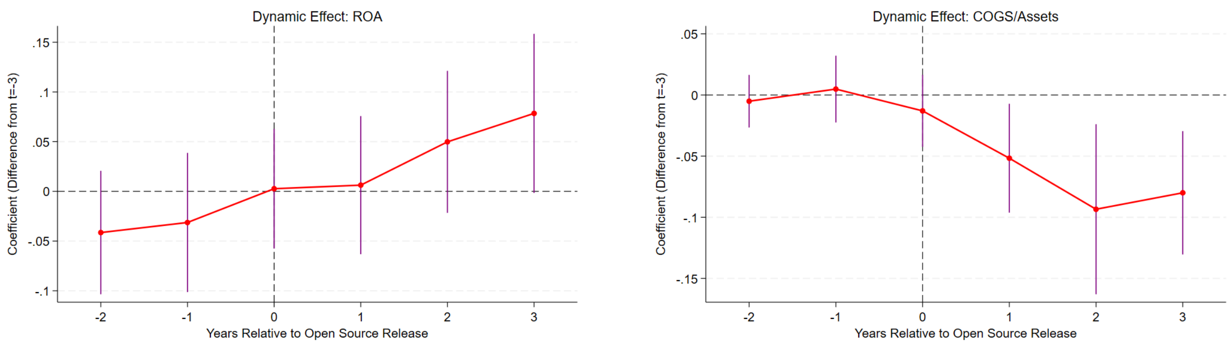



**Figure 3. Dynamic Treatment Effect on Peer Firm Sales**
This figure plots the dynamic treatment effect of focal firms' open-source releases on peer firms' Sales normalized by total assets. The estimation utilizes the full sample with the absolute cohort-level Kogan (2017) innovation value as an analytic weight. The specification includes Cohort-Firm and Cohort-Time fixed effects, firm-level controls (Market Capitalization, Market-to-Book Ratio, Leverage), and firm-level clustered standard errors. The x-axis indicates the years relative to the focal firm's open-source release (Year 0). The y-axis reports the estimated coefficients as the difference from the base year ($t = -3$), accompanied by their 95% confidence intervals.




**(a)** Return on Assets (ROA) **(b)** Cost of Goods Sold (COGS)

**Figure 4. Dynamic Treatment Effect on Peer Firm Profitability and Cost**
This figure presents the dynamic treatment effects on the profitability and cost measures of peer firms. Panel (a) plots the estimated effect on Return on Assets (ROA), and Panel (b) plots the effect on Cost of Goods Sold (COGS) normalized by total assets. For both panels, the x-axis indicates the years relative to the focal firm's open-source release (Year 0), and the y-axis shows the estimated coefficients relative to the base year ($t = -3$), accompanied by their 95% confidence intervals. All estimates are derived from the weighted DiD specification incorporating Cohort-Firm and Cohort-Time fixed effects, firm-level controls, and firm-level clustered standard errors.

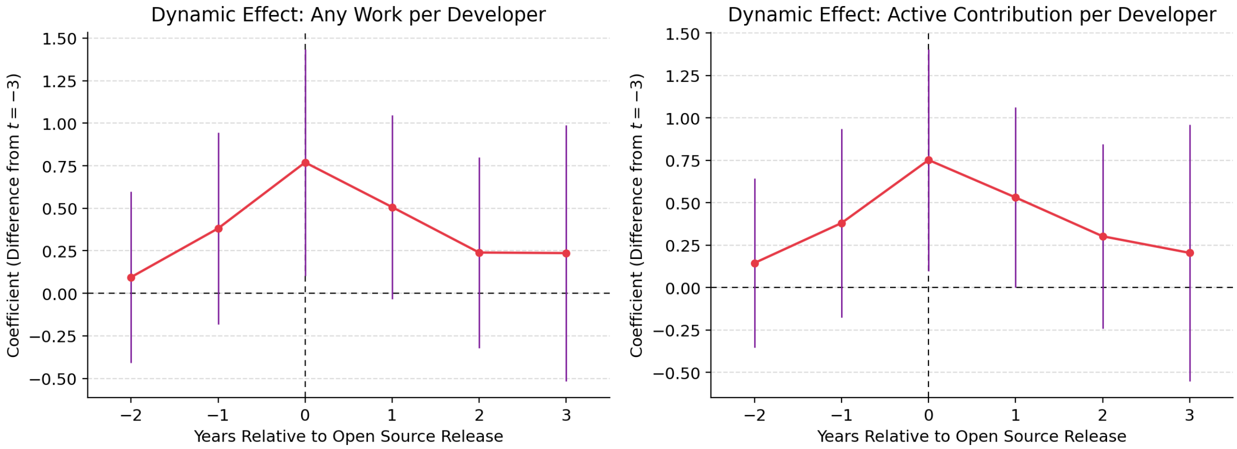



**Figure 5. Dynamic Effect on Peer Firm GitHub Activity per Developer**
This figure plots the dynamic treatment effect of focal firms' open-source releases on peer firms' per-developer GitHub activity. Panel (A) shows any contribution per working developer (pushes, pull requests, issues, issue comments, forks, stars, and watches); Panel (B) shows active contributions per working developer (pushes, pull requests, issues, issue comments, and forks). The estimation utilizes the Top 100 BigQuery-linked sub-sample with the cohort-level absolute Kogan (2017) innovation value as an analytic weight, matching Panel B of Table 5. The specification includes Cohort-Firm and Cohort-Time fixed effects, firm-level controls (Market Capitalization, Market-to-Book Ratio, Leverage), and firm-clustered standard errors. The x-axis indicates years relative to the focal firm's open-source release (Year 0); the y-axis reports the estimated coefficients as the difference from the base year ($t = -3$) with 95% confidence intervals. The pre-treatment coefficients are statistically indistinguishable from zero, supporting the parallel-trends assumption. The corresponding full-sample dynamic estimates are reported in Internet Appendix Figure IA1.

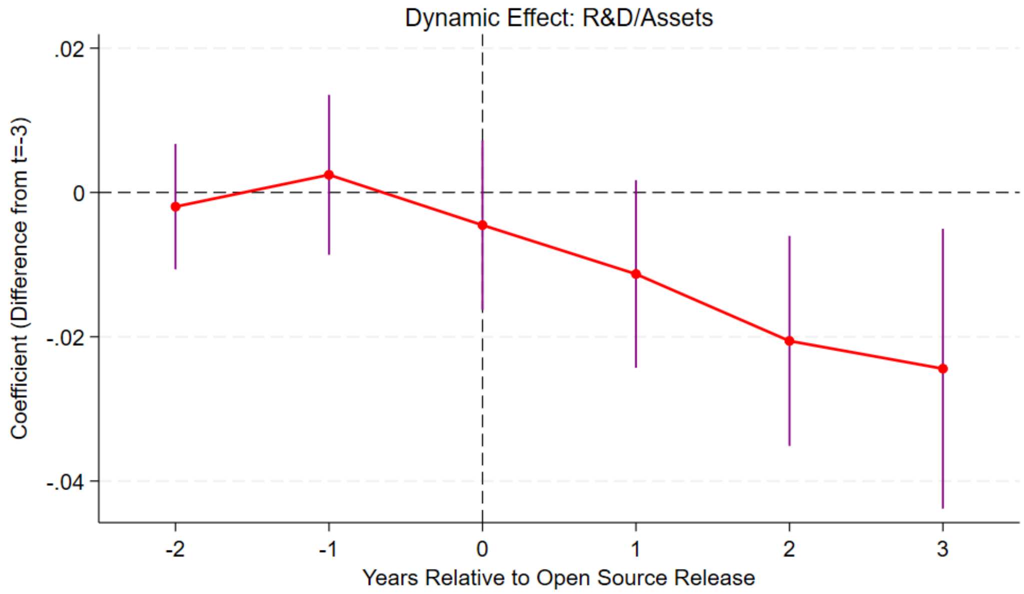



**Figure 6. Dynamic Treatment Effect on Peer Firm R&D**

This figure plots the dynamic treatment effect of focal firms' open-source releases on peer firms' Research and Development (R&D) expenditure normalized by total assets. The specification includes Cohort-Firm and Cohort-Time fixed effects, firm-level controls (Market Capitalization, Market-to-Book Ratio, Leverage), firm-level clustered standard errors, and the absolute cohort-level Kogan (2017) innovation value as the analytic weight. The x-axis indicates the years relative to the focal firm's open-source release (Year 0). The y-axis reports the estimated coefficients as the difference from the base year ($t = -3$), accompanied by their 95% confidence intervals.

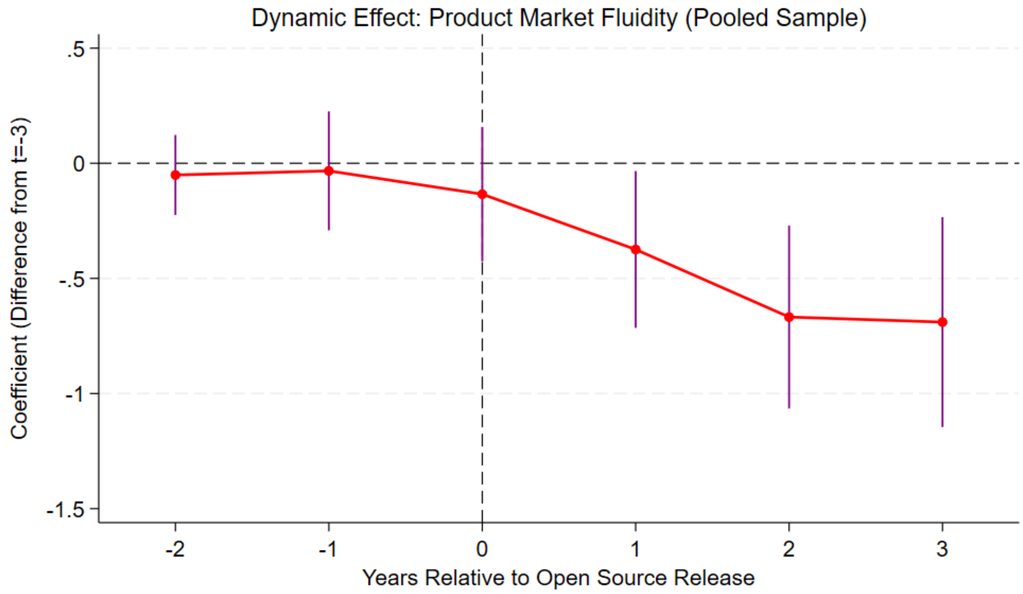
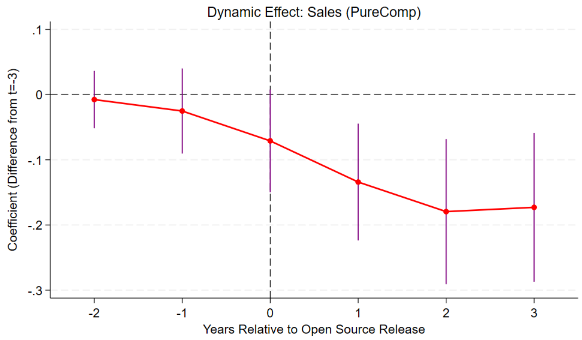
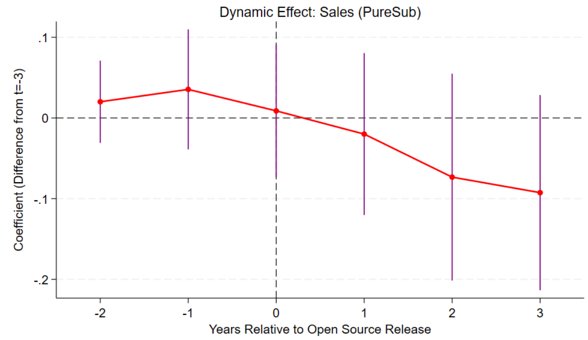



**Figure 7. Dynamic Treatment Effect on Peer Firm Product Market Fluidity**
This figure plots the dynamic treatment effect of focal firms' open-source releases on peer firms' Product Market Fluidity, a text-based measure capturing the intensity with which competitors reshape their product descriptions relative to a focal firm's product vocabulary. The specification follows the weighted DiD approach with Cohort-Firm and Cohort-Time fixed effects, firm-level controls, and firm-level clustered standard errors. The x-axis indicates the years relative to the focal firm's open-source release (Year 0). The y-axis reports the estimated coefficients as the difference from the base year ($t = -3$), accompanied by their 95% confidence intervals.




**(a)** Pure Complement Peers




**(b)** Pure Substitute Peers

**Figure 8. Heterogeneous Dynamic Effect on Sales: Complement vs. Substitute Peers**
This figure plots the heterogeneous dynamic treatment effects on Sales for peer firms, partitioned by CPC technology overlap (TF-IDF cosine similarity with the releasing repository) and industry-adjusted stock-return correlation with the focal firm around the event window. Panel (a) displays the effect on Pure Complement peers, defined by high technology overlap and positive return co-movement. Panel (b) displays the effect on Pure Substitute peers, defined by high technology overlap and negative return co-movement. The dynamic Difference-in-Differences estimates are weighted by the absolute cohort-level Kogan (2017) innovation value. The x-axis indicates years relative to the release (Year 0), and the y-axis reports the coefficients relative to the base year ($t = -3$), accompanied by 95% confidence intervals.

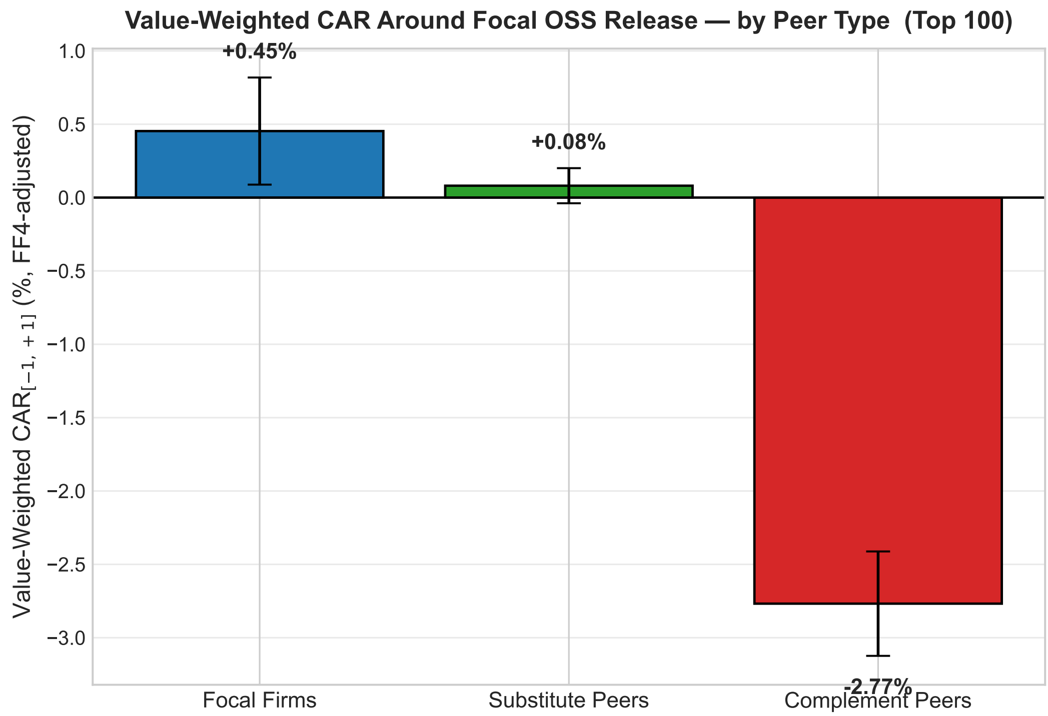



**Figure 9. Value-Weighted CAR$_{[-1,+1]}$ by Peer Type: Focal, Substitute, Complement**

This figure displays the value-weighted cumulative abnormal returns (CAR) over the $[-1, +1]$ event window around each open-source release, partitioned into focal firms, substitute peers, and complement peers. The sample is restricted to the top 100 strategically influential open-source releases. Abnormal returns are estimated using the Fama-French four-factor (FF4) model, with parameters fit on a 150-day estimation window ending 30 days prior to the event. Firms are weighted by their pre-event market capitalization. Error bars represent 95% confidence intervals, constructed from the effective-sample-size weighted cross-sectional variance. Substitutes and complements are partitioned by the sign of pre-event return co-movement with the focal firm among peers with high CPC technological overlap.

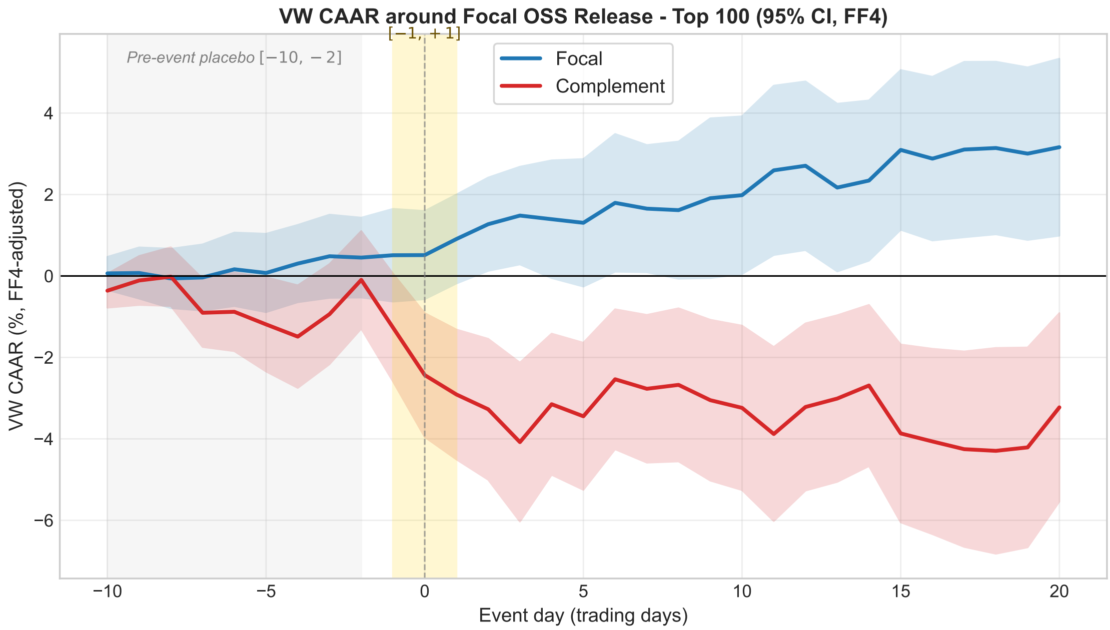



**Figure 10. Event-Window Dynamics of Value-Weighted CAAR: Focal vs. Complement Peers**

This figure traces the value-weighted cumulative average abnormal return (CAAR) of focal firms and complement peers from trading day −10 through day +20 relative to the open-source release (Day 0). The sample comprises the top 100 strategically influential releases. Abnormal returns are estimated from the Fama-French four-factor (FF4) model and weighted by pre-event market capitalization. Shaded bands denote 95% confidence intervals, constructed from the effective-sample-size weighted cross-sectional variance. The gray-shaded region marks the pre-event placebo window $[-10, -2]$, during which both groups exhibit flat pre-trends ($t = +0.90$ for focal, $t = -0.16$ for complement). The gold-shaded region marks the baseline event window $[-1, +1]$. Substitute peers are omitted from the figure for clarity because their entire trajectory is statistically indistinguishable from zero; they appear in the accompanying table and the full-sample robustness appendix.

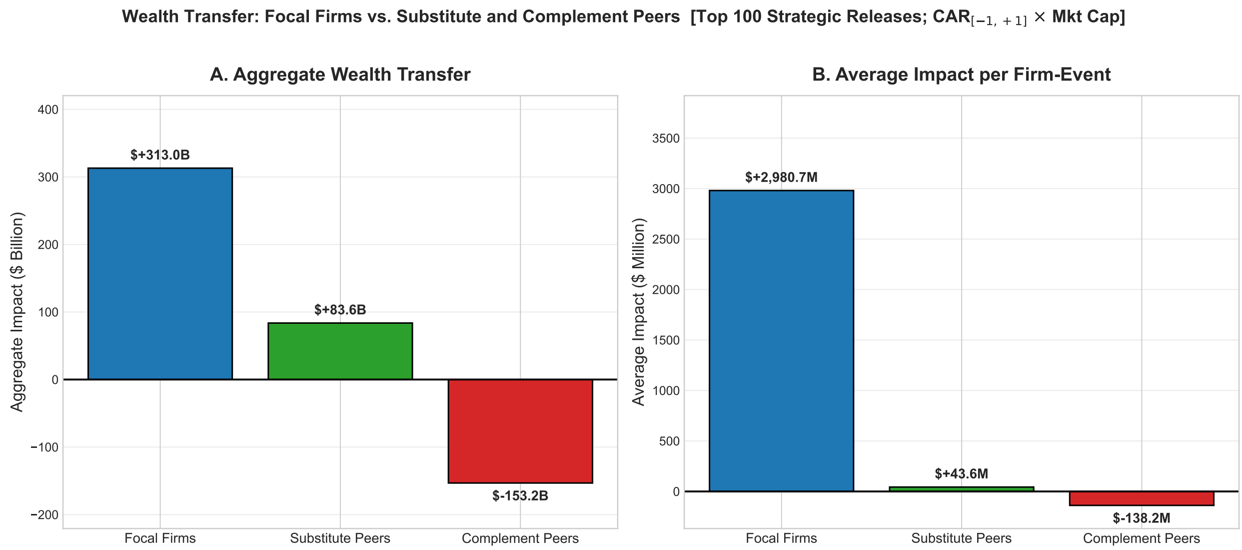



**Figure 11. Aggregate Market-Value Impact by Peer Type**

This figure illustrates the aggregate dollar-value impact of open-source software releases across focal firms, substitute peers, and complement peers. The dollar impact for each firm is computed as the Fama-French four-factor CAR over the $[-1, +1]$ event window multiplied by its market capitalization on the trading day immediately preceding the release, and then summed within each partition. Complement and substitute peers are defined by CPC technological overlap and the sign of pre-event return co-movement with the focal firm, as described in Section 6.1.

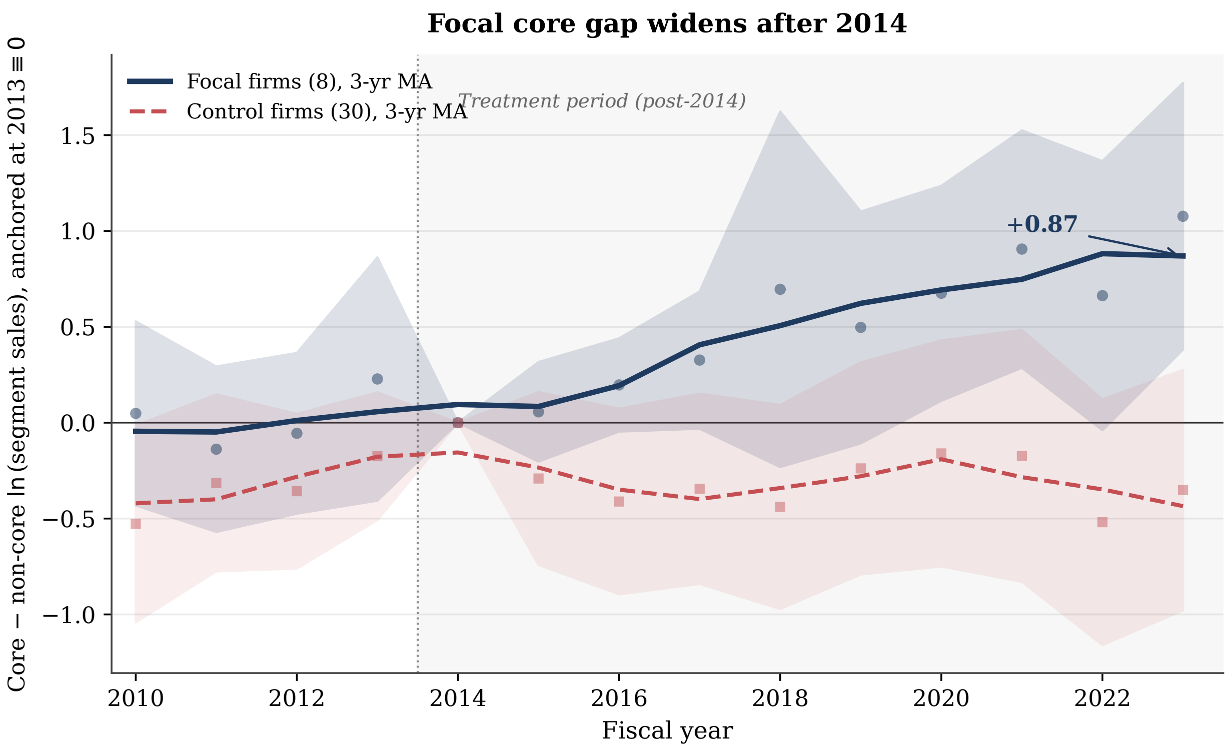



**Figure 12. Focal Firm Demand Response: Event-Time Dynamics**

This figure plots the core-minus-non-core difference in ln(segment sales) by fiscal year, separately for the 8 focal firms and the 30 non-releasing tech-firm controls, after absorbing firm×year and firm×segment fixed effects. Each series is a three-year moving average with 95% confidence bands, and the shaded region marks the post-2014 treatment period. The focal core gap is flat before 2014 and rises thereafter, while the control gap stays near zero, consistent with the cumulative-dose response in Table 10.

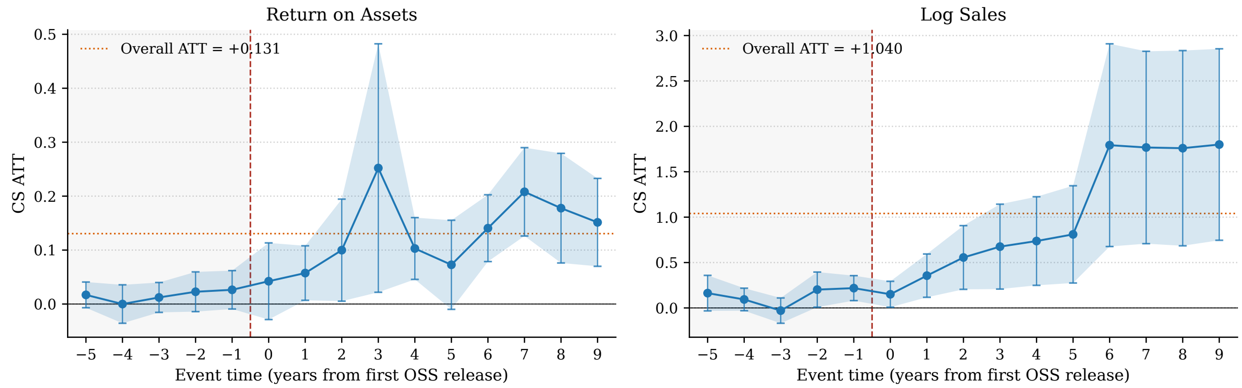



**Figure 13. Callaway–Sant'Anna Dynamic Event Study: ROA and Log Sales**
This figure plots the Callaway–Sant'Anna doubly-robust ATT for the two headline financial outcomes (Return on Assets in the left panel, Log Sales in the right panel) as a function of event time relative to the focal firm's open-source release. The estimator uses the never-treated firms as the control group and aggregates clean 2×2 cohort-time comparisons under the dynamic specification. Vertical bars are 95% pointwise confidence intervals; the orange dashed line in each panel reports the overall ATT across the post-treatment window. The pre-treatment coefficients lie within the confidence band around zero, supporting the parallel-trends assumption. The post-treatment trajectory is positive and monotonically increasing for both outcomes through year nine after the release, with ROA significant from event time +1 onward.

**Table 1. Descriptive Summary of the Full Open-Source Release Sample**

This table presents descriptive statistics for the 1,671 open-source releases utilized in the empirical analysis. Panel A reports summary statistics by quartiles of the Kogan (2017) innovation value, defined as the focal firm's market-value reaction (in millions of U.S. dollars) attributable to the release. The columns report the number of releases (*N*), the average cumulative abnormal return around the release window (CAR, %), the average intensity ($M), average GitHub Stars one year post-release, and the releasing firm's average market capitalization ($B) on the release date. Panel B reports the top 15 focal firms ranked by the number of releases, alongside their total intensity ($M), average CAR (%), and total GitHub Stars accumulated across their respective releases. The sample period spans from 2015 to 2024.

**Panel A: Summary by Kogan Intensity Quartiles**

| Quartile | N | Avg. CAR (%) | Avg. Intensity ($M) | Avg. Stars | Avg. Mkt Cap ($B) |
|---|---|---|---|---|---|
| Q1 (Lowest) | 427 | -2.05 | -16,857.4 | 5,472 | 975.6 |
| Q2 | 427 | -1.91 | -1,078.8 | 5,017 | 313.5 |
| Q3 | 427 | +0.96 | +2,681.1 | 7,880 | 417.4 |
| Q4 (Highest) | 426 | +2.46 | +19,763.4 | 5,225 | 1,003.6 |

**Panel B: Top 15 Focal Firms by Release Volume**

| Company Name | # of Repos | Total Intensity ($M) | Avg. CAR (%) | Total Stars |
|---|---|---|---|---|
| Microsoft | 534 | 1,105,058 | +0.30 | 2,966,059 |
| Alphabet | 469 | 48,991 | +0.09 | 3,435,776 |
| Meta | 291 | 139,906 | +0.31 | 2,220,525 |
| Amazon | 101 | -281,229 | -0.24 | 309,247 |
| Nvidia | 100 | 316,684 | +0.52 | 378,405 |
| Apple | 52 | 450,913 | +0.58 | 303,754 |
| Netflix | 43 | 1,725 | -0.01 | 208,381 |
| Salesforce | 22 | -96 | -26.09 | 54,539 |
| Uber | 19 | 0.3 | +0.42 | 3,049 |
| IBM | 16 | 1,446 | +0.06 | 31,865 |
| Oracle | 10 | 3,506 | +0.15 | 44,552 |
| Spotify | 9 | -2 | -3.75 | 0 |
| Adobe | 8 | -906 | +0.58 | 44,744 |
| Cisco | 6 | 2,827 | +0.34 | 20,536 |
| Atlassian | 6 | -53 | +0.86 | 9,548 |

**Table 2. Summary Statistics: Treated vs. Matched Control**

This table presents pre-event year ($t = -1$) summary statistics for the main firm-year level variables used in the empirical analysis. Panel A reports the statistics for the treated cohort, while Panel B reports the matched-control peer firms. Financial ratios (Sales, COGS, R&D) are normalized by total assets. Market-to-Book Ratio is the market value of equity divided by the book value of equity. Market Capitalization and Total Assets are expressed in billions of U.S. dollars. Product Market Fluidity, Product Market Scope, and Self Fluidity measure competitive pressure and product space dynamics. Counts (*N*) refer to cohort-firm observations. The sample period spans 2015–2024.

| | N | Mean | Std. Dev. | p10 | Median | p90 |
|---|---|---|---|---|---|---|
| **Panel A: Treated Sample** | | | | | | |
| Sales | 86,614 | 0.783 | 0.588 | 0.317 | 0.629 | 1.440 |
| COGS | 86,614 | 0.342 | 0.465 | 0.052 | 0.191 | 0.805 |
| R&D | 86,616 | 0.131 | 0.120 | 0.018 | 0.107 | 0.246 |
| ROA | 86,614 | -0.087 | 0.629 | -0.276 | -0.006 | 0.119 |
| Market Cap ($ Billions) | 86,616 | 32.293 | 131.185 | 0.091 | 2.388 | 47.320 |
| Market-to-Book | 86,616 | 5.516 | 122.736 | 1.074 | 4.379 | 19.202 |
| Leverage | 86,616 | 0.192 | 0.194 | 0.000 | 0.149 | 0.450 |
| Total Assets ($ Billions) | 86,616 | 12.895 | 44.720 | 0.065 | 0.965 | 18.769 |
| Product Market Fluidity | 78,229 | 5.824 | 2.006 | 3.401 | 5.741 | 8.035 |
| Self Fluidity | 70,891 | 19.030 | 14.263 | 5.804 | 15.190 | 36.973 |
| Product Market Scope | 76,956 | 15.042 | 8.160 | 5.000 | 14.000 | 27.000 |
| **Panel B: Matched Control Sample** | | | | | | |
| Sales | 66,668 | 0.664 | 0.430 | 0.246 | 0.584 | 1.132 |
| COGS | 66,668 | 0.307 | 0.340 | 0.050 | 0.184 | 0.712 |
| R&D | 66,668 | 0.080 | 0.087 | 0.000 | 0.058 | 0.188 |
| ROA | 66,668 | -0.071 | 0.463 | -0.293 | 0.010 | 0.126 |
| Market Cap ($ Billions) | 66,668 | 23.304 | 126.379 | 0.068 | 2.007 | 30.297 |
| Market-to-Book | 66,668 | 6.555 | 59.683 | 0.882 | 3.614 | 16.479 |
| Leverage | 66,668 | 0.201 | 0.251 | 0.000 | 0.140 | 0.458 |
| Total Assets ($ Billions) | 66,668 | 8.355 | 31.907 | 0.060 | 1.033 | 14.220 |
| Product Market Fluidity | 41,184 | 5.215 | 2.211 | 2.834 | 5.004 | 7.613 |
| Self Fluidity | 36,991 | 21.016 | 17.874 | 5.471 | 15.446 | 42.687 |
| Product Market Scope | 41,298 | 10.279 | 7.088 | 2.000 | 9.000 | 21.000 |

**Table 3. The Impact of Big Tech Open-Source Release on Peer Firms' Sales**

This table presents the results of a stacked Difference-in-Differences (DiD) estimation of the effect of focal firms' open-source releases on the sales of peer firms. Column (1) presents a baseline OLS specification without fixed effects or controls. Column (2) incorporates Cohort-Firm and Cohort-Time fixed effects along with firm-level controls. Column (3) further weights the observations using the absolute value of the cohort-level Kogan (2017) innovation value to address potential look-ahead bias. Control variables include Market Capitalization, Market-to-Book (M/B) Ratio, and Leverage. Standard errors are clustered at the firm level in Columns (2) and (3). $t$-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans from 2015 to 2024.

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable = | Sales | Sales | Sales |
| Treated × Post | -0.049*** | -0.046** | -0.060*** |
| | (-21.58) | (-2.39) | (-2.69) |
| Market Cap | | 0.091*** | 0.081** |
| | | (2.69) | (2.28) |
| M/B Ratio | | -0.002 | -0.002 |
| | | (-0.88) | (-1.51) |
| Leverage | | 0.069* | 0.071** |
| | | (1.91) | (2.07) |
| Adjusted $R^2$ | 0.009 | 0.715 | 0.735 |
| Cohort-Firm FE | No | Yes | Yes |
| Cohort-Time FE | No | Yes | Yes |
| Clustering | None | Firm | Firm |
| Weight | No | No | Yes |
| Mean of Dep. Var. | 0.728 | 0.728 | 0.728 |
| Observations | 974,473 | 974,277 | 974,277 |

**Table 4. The Impact of Big Tech Open-Source Release on Peer Firms' Profitability and Costs**
This table presents the results of a stacked Difference-in-Differences (DiD) estimation of the effect of focal firms' open-source releases on the profitability and costs of peer firms. The dependent variables are Return on Assets (ROA) and Cost of Goods Sold (COGS), both normalized by total assets. Columns (1) and (3) present a baseline OLS specification without fixed effects or controls. Columns (2) and (4) incorporate Cohort-Firm and Cohort-Time fixed effects along with firm-level controls, and weight the observations using the absolute value of the cohort-level Kogan (2017) innovation value to address potential look-ahead bias. Control variables include Market Capitalization, Market-to-Book (M/B) Ratio, and Leverage. Standard errors are clustered at the firm level in Columns (2) and (4). $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans from 2015 to 2024.

| | Profitability | | Cost | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Dependent Variable = | ROA | ROA | COGS | COGS |
| Treated × Post | 0.018$^{***}$ | 0.053$^{**}$ | -0.045$^{***}$ | -0.054$^{***}$ |
| | (7.46) | (2.00) | (-25.58) | (-3.00) |
| Market Cap | | -0.014 | | 0.017 |
| | | (-0.36) | | (1.46) |
| M/B Ratio | | -0.000 | | -0.002$^{*}$ |
| | | (-0.10) | | (-1.72) |
| Leverage | | -1.066$^{***}$ | | 0.034 |
| | | (-18.00) | | (0.98) |
| Adjusted $R^2$ | 0.000 | 0.589 | 0.001 | 0.759 |
| Cohort-Firm FE | No | Yes | No | Yes |
| Cohort-Time FE | No | Yes | No | Yes |
| Clustering | None | Firm | None | Firm |
| Weight | No | Yes | No | Yes |
| Mean of Dep. Var. | -0.064 | -0.064 | 0.328 | 0.328 |
| Observations | 974,473 | 974,277 | 974,473 | 974,277 |

**Table 5. The Impact of Big Tech Open-Source Releases on Peer Firms' GitHub Activity**

This table presents stacked Difference-in-Differences (DiD) estimates of the effect of focal firms' open-source releases on the GitHub engagement of peer firms. Both panels weight observations by the cohort-level absolute Kogan (2017) innovation value to anchor the estimates on the most economically influential releases. Panel A reports the full sample of approximately 1,700 Kogan-eligible cohorts. Panel B restricts the analysis to the Top 100 BigQuery-linked sub-sample, in which contributor identity is most precisely linked to firm-level employment; this is the identification sample matching Figure 5. Columns (1) to (3) report results for per-developer activity (total, active, and passive contributions); column (4) reports the effect on the size of the active developer base. *Active* contributions include pushes, pull requests, issues, issue comments, and forks; *Passive* contributions include stars and watches. All specifications include firm-level controls (Market Capitalization, Market-to-Book Ratio, Leverage), Cohort-Firm and Cohort-Time fixed effects, and firm-clustered standard errors. Cohort × year fixed effects absorb secular GitHub-adoption trends. Unweighted variants of both panels are reported in Internet Appendix Table IA.7. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans 2015–2024.

| Dependent Variable = | Activity per Developer | | | Developer Base |
|---|---|---|---|---|
| | (1) Any Work per Dev. | (2) Active Contrib. | (3) Passive Contrib. | (4) Ln(Working Devs + 1) |
| *Panel A: Full sample* | | | | |
| Treated $\times$ Post | 1.145$^{***}$ | 1.134$^{**}$ | 0.011 | 0.141$^{***}$ |
| | (2.60) | (2.57) | (0.41) | (5.32) |
| Adjusted $R^2$ | 0.743 | 0.740 | 0.422 | 0.839 |
| Mean of Dep. Var. | 3.122 | 2.801 | 0.321 | 0.480 |
| Observations | 974,385 | 974,385 | 974,385 | 974,385 |
| *Panel B: Top 100 sub-sample* | | | | |
| Treated $\times$ Post | 0.347$^{**}$ | 0.341$^{**}$ | 0.006 | 0.070 |
| | (2.17) | (2.19) | (0.23) | (1.47) |
| Adjusted $R^2$ | 0.137 | 0.093 | 0.343 | 0.749 |
| Mean of Dep. Var. | 1.087 | 0.765 | 0.322 | 0.606 |
| Observations | 33,834 | 33,834 | 33,834 | 33,834 |
| Controls | Yes | Yes | Yes | Yes |
| Cohort-Firm FE | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes |
| Clustering | Firm | Firm | Firm | Firm |
| Weight | Kogan $\|w\|$ | Kogan $\|w\|$ | Kogan $\|w\|$ | Kogan $\|w\|$ |

**Table 6. The Impact of Big Tech Open-Source Releases on Peer Firms' Innovation**

This table presents the results of a stacked Difference-in-Differences (DiD) estimation of the effect of focal firms' open-source releases on the innovation outcomes of peer firms. The dependent variables are R&D expenditure (normalized by total assets), average patent portfolio age, and average exploratory patent age. Columns (1), (3), and (5) present a baseline OLS specification without fixed effects or controls. Columns (2), (4), and (6) incorporate Cohort-Firm and Cohort-Time fixed effects along with firm-level controls, and weight the observations using the absolute value of the cohort-level Kogan (2017) innovation value to address potential look-ahead bias. Control variables include Market Capitalization, Market-to-Book (M/B) Ratio, and Leverage. Standard errors are clustered at the firm level in the full specifications. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans from 2015 to 2024.

| Dependent Variable = | R&D | | Patent Pf. Age | | Exploratory Pf. Age | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated × Post | -0.006$^{***}$ | -0.014$^{***}$ | 0.476$^{***}$ | 0.323$^{***}$ | 0.744$^{***}$ | 0.792$^{***}$ |
| | (-11.64) | (-2.77) | (21.11) | (2.84) | (27.86) | (5.19) |
| Market Cap | | 0.013$^{**}$ | | -0.074 | | 0.771$^{***}$ |
| | | (2.11) | | (-0.51) | | (2.99) |
| M/B Ratio | | 0.001$^{**}$ | | 0.004 | | 0.011 |
| | | (2.48) | | (1.12) | | (1.32) |
| Leverage | | 0.262$^{***}$ | | 0.019 | | 0.053 |
| | | (7.36) | | (0.54) | | (1.56) |
| Adjusted $R^2$ | 0.029 | 0.788 | 0.023 | 0.938 | 0.052 | 0.954 |
| Cohort-Firm FE | No | Yes | No | Yes | No | Yes |
| Cohort-Time FE | No | Yes | No | Yes | No | Yes |
| Clustering | None | Firm | None | Firm | None | Firm |
| Weight | No | Yes | No | Yes | No | Yes |
| Mean of Dep. Var. | 0.107 | 0.107 | 5.192 | 5.195 | 6.146 | 6.148 |
| Observations | 974,582 | 974,385 | 974,895 | 974,385 | 974,895 | 974,385 |

**Table 7. The Impact of Big Tech Open-Source Releases on Peer Firms' Product Space**

This table presents the results of a stacked Difference-in-Differences (DiD) estimation of the effect of focal firms' open-source releases on the product-space dynamics of peer firms. The dependent variables are Product Market Fluidity and Product Market Scope (based on Doc2Vec). Columns (1) and (3) present a baseline OLS specification without fixed effects or controls. Columns (2) and (4) incorporate Cohort-Firm and Cohort-Time fixed effects along with firm-level controls, and weight the observations using the absolute value of the cohort-level Kogan (2017) innovation value to address potential look-ahead bias. Control variables include Market Capitalization, Market-to-Book (M/B) Ratio, and Leverage. Standard errors are clustered at the firm level in the full specifications. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans from 2015 to 2024.

| Dependent Variable = | Product Market Fluidity | | Product Market Scope | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated × Post | -0.515$^{***}$ | -0.374$^{***}$ | -0.028 | 0.341 |
| | (-51.34) | (-3.07) | (-0.72) | (0.71) |
| Market Cap | | 0.140 | | 2.354$^{**}$ |
| | | (0.67) | | (2.27) |
| M/B Ratio | | 0.018$^{**}$ | | -0.044 |
| | | (2.07) | | (-1.08) |
| Leverage | | -0.139$^{***}$ | | 0.105 |
| | | (-3.25) | | (0.64) |
| Adjusted $R^2$ | 0.041 | 0.578 | 0.061 | 0.789 |
| Cohort-Firm FE | No | Yes | No | Yes |
| Cohort-Time FE | No | Yes | No | Yes |
| Clustering | None | Firm | None | Firm |
| Weight | No | Yes | No | Yes |
| Mean of Dep. Var. | 5.334 | 5.334 | 13.282 | 13.284 |
| Observations | 798,325 | 797,806 | 739,197 | 736,284 |

**Table 8. Heterogeneous Effects on Substitute vs. Complement Peers (CPC Classification)**

This table reports stacked Difference-in-Differences (DiD) estimates of the effect of focal firms' open-source releases on peer firms' Sales and Cost of Goods Sold (COGS), separately for Substitute and Complement peers classified by CPC technology overlap. A peer is labeled a Substitute if its patent portfolio's CPC cosine similarity with the releasing repository exceeds the cohort median and its industry-adjusted stock-return correlation around the release window is negative; Complements are peers with above-median CPC overlap and positive return co-movement. The dependent variables are normalized by total assets. All specifications include Cohort-Firm and Cohort-Time fixed effects and firm-level controls, and observations are weighted by the absolute value of the cohort-level Kogan (2017) innovation value. Standard errors are clustered at the firm level. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans from 2015 to 2024.

| Dependent Variable = | Sales | | COGS | |
|---|---|---|---|---|
| Peer Firm Type = | (1) Substitute | (2) Complement | (3) Substitute | (4) Complement |
| Treated × Post | -0.0335 | -0.0928$^{***}$ | -0.0376$^{**}$ | -0.0643$^{***}$ |
| | (-1.36) | (-3.20) | (-1.97) | (-2.79) |
| Adjusted $R^2$ | 0.654 | 0.725 | 0.661 | 0.747 |
| Controls | Yes | Yes | Yes | Yes |
| Cohort-Firm FE | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes |
| Clustering | Firm | Firm | Firm | Firm |
| Weight | Yes | Yes | Yes | Yes |
| Mean of Dep. Var. | 0.666 | 0.729 | 0.288 | 0.352 |
| Observations | 554,830 | 559,006 | 554,830 | 559,006 |

**Table 9. Value-Weighted Cumulative Abnormal Returns by Peer Type**

This table reports value-weighted cumulative abnormal returns (CAR) around the announcement dates of the top 100 strategically influential open-source releases, partitioned into focal firms, substitute peers, and complement peers. Abnormal returns are estimated using the Fama-French four-factor (FF4) model, with factor loadings fit on a 150-day estimation window ending 30 days prior to the event. Firms are weighted by their pre-event market capitalization. The window $[-10,-2]$ in column (1) serves as a pre-event placebo window and is required to be statistically flat for a clean identification. Columns (2)–(5) report post-event windows of increasing length starting from the day before the release. Substitutes and complements are partitioned by the sign of pre-event return co-movement with the focal firm, conditional on high CPC technological overlap, as described in Section 6.1. $t$-statistics (in parentheses) are computed from the effective-sample-size weighted cross-sectional variance. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

| Group | Metric | (1) **[–10,–2]** (Placebo) | (2) **[–1,+1]** (Main) | (3) **[–1,+3]** | (4) **[–1,+10]** | (5) **[–1,+20]** |
|---|---|---|---|---|---|---|
| **Focal** | Mean CAR | +0.45% | +0.46%* | +1.03%*** | +1.53%* | +2.71%** |
| (N=67) | $t$ | (+0.90) | (+1.65) | (+2.83) | (+1.65) | (+2.48) |
| **Substitute** | Mean CAR | −0.18% | +0.08% | −0.14% | −0.63% | −0.94% |
| (N=842) | $t$ | (−0.44) | (+0.21) | (−0.28) | (−1.01) | (−1.15) |
| **Complement** | Mean CAR | −0.10% | −2.81%*** | −3.98%*** | −3.14%*** | −3.13%*** |
| (N=1,096) | $t$ | (−0.16) | (−6.43) | (−5.99) | (−4.41) | (−3.51) |

**Table 10. Focal Firm Demand Response to Open-Source Release: Dose-Response Ladder**

This table reports a four-step dose-response ladder of the focal-firm demand response to strategic open-source release. The outcome is ln(segment sales), taken from Compustat Segment data over fiscal years 2010–2023. "Core" segments are hand-mapped from each focal firm's 10-K business-segment disclosures (Microsoft Intelligent Cloud, Alphabet Google Cloud, Adobe Digital Media, Nvidia Compute & Networking, and so on); "non-core" segments are the remaining business segments of the same firm. All specifications use a combined panel of 8 focal firms plus 30 non-releasing large tech firms (Compustat SIC 7370–7379 or NAICS 5112/5415/5418, sale $\geq$ \$2B for at least five years, excluding the 8 focals); control firms have cumulative $|\kappa| = 0$ by construction and anchor the slope identification. The treatment intensity is the firm-year cumulative Kogan intensity $\sum_{r:\text{year}(r)\leq t} |\kappa_r|$, scaled to trillion-dollar units. Column (1) reports the baseline contemporaneous specification. Column (2) replaces the contemporaneous dose with its one-year lag, $\sum_{r:\text{year}(r)\leq t-1} |\kappa_r|$, allowing demand to absorb releases with the lag that segment sales realization typically requires. Columns (3) and (4) restrict the sample to fiscal years 2014–2023, the period during which strategic open-source release became a first-order corporate activity at the focal firms, repeating the contemporaneous and lagged specifications on this treatment-period sample. All specifications include firm$\times$year and firm$\times$segment fixed effects, so that firm-wide year shocks (e.g., the post-2015 cloud boom) and segment-baseline levels are absorbed. Standard errors are clustered at the firm$\times$segment level; $t$-statistics in parentheses. $^{***}$, $^{**}$, and $^{*}$ denote statistical significance at the 1%, 5%, and 10% levels.

| | *Dependent variable:* ln(segment sales) | | | |
|---|---|---|---|---|
| | (1) Baseline | (2) Lagged (one-year lag) | (3) Post-2014 baseline | (4) Post-2014 + lagged |
| *IsCore* $\times$ $\lvert\kappa\rvert_t^{\text{cum}}$ (per \$1T) | $+0.082^{***}$ | – | $+0.107^{***}$ | – |
| | (+2.74) | | (+4.85) | |
| *IsCore* $\times$ $\lvert\kappa\rvert_{t-1}^{\text{cum}}$ (per \$1T) | – | $+0.093^{***}$ | – | $+0.120^{***}$ |
| | | (+2.86) | | (+4.70) |
| Sample period | 2010–2023 | 2010–2023 | 2014–2023 | 2014–2023 |
| Firms | 38 | 38 | 38 | 38 |
| Segment-years | 1,925 | 1,925 | 1,333 | 1,333 |
| Firm $\times$ Year FE | Yes | Yes | Yes | Yes |
| Firm $\times$ Segment FE | Yes | Yes | Yes | Yes |
| Cluster | Firm $\times$ Segment | | | |

**Table 11. Callaway–Sant'Anna Robustness: Headline Outcomes**

This table addresses the heterogeneous-treatment-effects concern raised by Goodman-Bacon (2021) and Sun and Abraham (2021). With staggered treatment timing, the two-way fixed-effect (TWFE) DiD estimator can use already-treated units as implicit controls for later-treated cohorts, generating forbidden comparisons that may attenuate estimated treatment effects. I re-estimate the focal-firm DiD using the Callaway and Sant'Anna (2021) doubly-robust ATT estimator with a never-treated control group, which weights only clean $2\times2$ cohort-time comparisons. The table reports the two headline financial outcomes (Return on Assets and log Sales). For each outcome, columns (1)–(2) report the TWFE coefficient $\hat{\beta}$ from the saturated specification with cohort-firm and cohort-time fixed effects; columns (3)–(4) report the Callaway–Sant'Anna overall ATT and its analytical standard error. Sign is preserved on both outcomes; CS estimates are larger and tighter than TWFE, indicating that the paper's main estimates are a conservative lower bound. Standard errors are clustered at the firm level for TWFE. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels.

| | Two-Way Fixed Effects | | Callaway–Sant'Anna | |
|---|---|---|---|---|
| Dependent Variable | $\hat{\beta}$ | $t$ | ATT | $t$ |
| Return on Assets | $+0.052^{**}$ | (2.20) | $+0.131^{***}$ | (4.95) |
| Log Sales | $+0.431^{*}$ | (1.89) | $+1.040^{***}$ | (3.94) |
| Firm FE | Yes | | — | |
| Year FE | Yes | | — | |
| Control Group | Full Panel | | Never-Treated | |
| Estimator | OLS, Cluster by Firm | | Doubly Robust | |
| Treated Firms | 8 | | 8 | |
| Control Firms | 443 | | 443 | |
| Observations | 3,406 | | 3,406 | |

# Internet Appendix for:
# "Commoditize Your Complement: Strategic Open-Source Releases and the Cannibalization of Rival Firms"

## A Theoretical Framework: The Multi-Segment Model

This section builds a framework in which strategic open-source release is the commoditization of one segment of a multi-segment firm's product portfolio. The construction proceeds from four primitives: the focal firm's three-segment profit structure, the open-source license as a commitment device, the peer firm's exposure vector across the same three segments, and four channels through which the release transmits to peer outcomes. From these primitives the framework generates six predictions. The first concerns the focal firm's own valuation. The second, which is the paper's centerpiece, is an asymmetric sign prediction across peer types that no alternative rationalization of the release reproduces. The remaining four characterize the mechanism chain through which peer outcomes adjust on real-economy margins.

### A.1 A Multi-Segment Firm as the Primitive Unit of Analysis

The analytical object is not the firm but the multi-segment firm, operating across three logically distinct product markets indexed by $j \in \{A, B, C\}$. Each segment has its own demand curve, its own set of rivals, and its own strategic role in the portfolio. Segment $A$ is the profit center, generating the monopoly or quasi-monopoly rents that the firm has every incentive to protect. Advertising inventory for Meta, search advertising for Alphabet, and enterprise cloud for Microsoft are canonical examples. Segment $B$ is a substitute market to $A$, where entry by a rival threatens the $A$-segment boundary; the focal firm wants a moat here as well. Segment $C$ is a complement market to $A$, in which higher consumption raises the marginal value of $A$ through a cross-market demand linkage. In segment $C$ the optimal posture is commoditization rather than rent extraction, because cheaper $C$-goods expand $A$-market demand. This three-segment partition restates the distinction that Adner and Kapoor (2010) draw between upstream components and downstream complements in innovation ecosystems, applied here to the firm's own internal portfolio rather than to an ecosystem of separate producers.

Let focal firm $f$ earn profit

$$\Pi_f = \pi_{f,A}(p_A, Q_C) + \pi_{f,B}(p_B) + \pi_{f,C}(p_C, p_{-C}), \tag{A-2}$$

where $p_j$ is the price in segment $j$, $Q_j$ is aggregate quantity, and $p_{-C}$ is the vector of rival $C$-

segment prices. The central assumption is

$$\frac{\partial \pi_{f,A}}{\partial Q_C} > 0. \tag{A1}$$

Higher consumption of $C$-goods expands the rent base of segment $A$. Assumption (A1) is the economic content of the commoditize-your-complement heuristic articulated informally by Spolsky (2002) and modeled formally by Parker and Van Alstyne (2005) in the two-sided network setting.

Peer firm $r$ is characterized by an exposure vector $(s_{r,A}, s_{r,B}, s_{r,C}) \in [0,1]^3$ with $\sum_j s_{r,j} = 1$, representing the fraction of the peer's revenue derived from each segment. A peer is not defined by an external label but by where its revenue lives. This representation makes the substitute-versus-complement distinction that organizes the empirical results an observable projection of the peer's segment exposure rather than an analyst-imposed taxonomy, a property we return to in Section A.3.

## A.2 Strategic Release as Commoditization of Segment $C$

The focal firm faces a discrete choice. It either extracts rent in segment $C$ through proprietary pricing or releases a high-quality asset $\Omega$ in segment $C$ under an open-source license at consumer price zero in perpetuity, financed out of segment $A$'s rents. Denote this choice $OSS_f \in \{0,1\}$.

Three features of the open-source license separate this choice from an ordinary price cut. First, the release is irreversible: once code is distributed under an open-source license, no actor including the releasing firm can return it to proprietary status, because any attempted closure is defeated by forks of the released version. Second, the subsidy flows across segments rather than within. The cost of providing $\Omega$ is absorbed by segment $A$ profits, not by an expectation that segment $C$ prices will eventually rise to recoup. Third, the release is a technical standard-setting act in addition to a pricing act. Adoption of $\Omega$ as common infrastructure creates network effects and switching costs that persist beyond the release announcement. The combination of these features makes the standard predatory-pricing hold-up logic inapplicable. A competitor cannot wait out a rival whose subsidy never exhausts because the subsidy does not come from the same pool the competitor is defending. This commitment structure, inherited from the legal form of open-source licensing rather than from discretionary managerial behavior, is what separates the strategy from an ordinary price war.

The change in focal profit from releasing $\Omega$ decomposes into four channels:

$$\Delta\Pi_f = \underbrace{\frac{\partial \pi_{f,A}}{\partial Q_C}\frac{\partial Q_C}{\partial OSS_f}}_{\text{(i) cross-market subsidy}} - \underbrace{\left|\frac{\partial \pi_{f,C}}{\partial OSS_f}\right|}_{\text{own } C\text{-rent loss}} + \underbrace{\mathbb{E}\left[\sum_r \Delta\pi_{r,C} \cdot \phi_{fr}\right]}_{\text{(ii) peer redirection}} + \underbrace{\lambda_3}_{\text{(iii) standards}} + \underbrace{\lambda_4}_{\text{(iv) talent}} . \tag{A-3}$$

Channel (i) is the cross-market subsidy gain that operates through assumption (A1). Channel (ii) captures the competitive consequence for rivals whose $C$-segment overlap with the focal firm, denoted $\phi_{fr}$, is non-zero; the redirection of rival activity accrues to the focal firm through downstream complementarities. Channel (iii), worth $\lambda_3 \geq 0$, is the option value of making $\Omega$ the technical standard, which in the tradition of Farrell and Saloner (1985) and Katz and Shapiro (1985) locks future entrants onto the focal firm's interface. Channel (iv), worth $\lambda_4 \geq 0$, is the option value of drawing outside developer contribution onto the focal firm's technology trajectory. Release is optimal when the sum of the four positive channels dominates the direct $C$-rent loss.

**Proposition 1 (Release under $A$-dominance):** *If $\pi_{f,A}$ is large relative to $\pi_{f,C}$ and $\partial\pi_{f,A}/\partial Q_C$ is sufficiently positive, then $OSS_f = 1$ is optimal even when $\partial\pi_{f,C}/\partial OSS_f < 0$.*

Proposition 1 explains why the set of firms observed to release strategic open-source software is concentrated among technology incumbents whose profit base is dominated by a protected $A$-segment. Microsoft, Alphabet, Meta, Apple, and Nvidia have the subsidy pool to finance a permanent giveaway in an adjacent complement market. A firm whose profit base is itself concentrated in segment $C$ has no equivalent pool from which to finance the release, and so does not release.

A corollary concerns the focal firm's announcement-window valuation. When the equity market observes the release and updates its expectation of $f$'s future profits, the positive contribution from channels (i), (iii), and (iv) dominates the negative own-rent contribution whenever the release is announced at all, because a release with negative net present value would not have been announced. The announcement return on the focal firm's equity is therefore non-negative in expectation.

**Hypothesis 1 (Focal-Firm Valuation Gain):** *Announcement of a strategic open-source release by a focal firm generates a non-negative cumulative abnormal return on the focal firm's equity in a narrow window around the release date.*

## A.3 Peer Classification and the Asymmetric Response Prediction

The framework's sharpest prediction concerns the cross-section of peer firms. Unlike the focal firm's valuation, which is bounded below by zero in the neighborhood of the release, peer valuations can take either sign, and the distribution of outcomes is pinned down by how each peer's revenue is composed over the three segments.

Consider peer $r$ with exposure vector $(s_{r,A}, s_{r,B}, s_{r,C})$. Define two statistics that summarize the peer's economic relation to the focal firm:

$$\sigma_r = s_{r,A}, \tag{A-4}$$

$$\gamma_r = s_{r,C}. \tag{A-5}$$

A peer whose revenue is concentrated in segment $A$ competes with the focal firm in the focal firm's own profit center. A peer whose revenue is concentrated in segment $C$ supplies products that complement the focal firm's $A$-segment, operating in the same segment that $\Omega$ commoditizes. Adopting the terminology used throughout the empirical section, a peer with high $\sigma_r$ and low $\gamma_r$ is a *substitute peer* to the focal firm, because its products substitute for the focal firm's core offering. A peer with high $\gamma_r$ and low $\sigma_r$ is a *complement peer* to the focal firm, because its products sit alongside the focal firm's core offering in the consumption bundle. Oracle and Salesforce, whose enterprise software competes with Microsoft's Dynamics and Azure services, are substitute-peer examples. MongoDB and Confluent, whose data-layer products run on top of the hyperscalers' core infrastructure, are complement-peer examples.

For a complement peer, the release directly compresses revenue. The peer sells in segment $C$, the same segment $\Omega$ now occupies at price zero. Consumers facing an $\Omega$-quality free alternative and the peer's proprietary alternative at price $p_{r,C}$ with switching cost $\tau$ adopt $\Omega$ whenever $V_\Omega \geq V_{r,C} - p_{r,C} - \tau$ (Zhu and Zhou, 2012). Rising $V_\Omega$ through successive focal-firm updates compresses $p_{r,C}$ toward marginal cost, and the complement peer's $C$-segment revenue base contracts. Because $\sigma_r \approx 0$, the complement peer has no offsetting revenue in focal's $A$-adjacent markets. Its market valuation falls as the discounted present value of the revenue contraction.

For a substitute peer, the release produces no direct $C$-segment compression because $\gamma_r \approx 0$. The $Q_C$ expansion induced by $\Omega$ raises the marginal value of segment $A$ through the cross-partial in assumption (A1), which operates for the substitute peer's own $A$-segment revenue exactly as it does for the focal firm's. This effect is positive in sign, though typically modest in magnitude, because the substitute peer also competes with the focal firm in $A$ and the focal firm's own expansion absorbs part of the $Q_C$-induced demand. The substitute peer's market valuation therefore adjusts by an amount that is weakly positive or statistically indistinguishable from zero.

**Proposition 2 (Asymmetric Response):** *Announcement of $OSS_f = 1$ generates a strongly negative cumulative abnormal return on complement peers and a weakly positive or null cumulative abnormal return on substitute peers. The magnitude of the complement-peer loss is increasing in $\gamma_r$; the magnitude of the substitute-peer gain is increasing in $\sigma_r$ and is bounded above by focal-firm absorption in segment $A$.*

No alternative rationalization of the release produces this asymmetry. A signaling account in which $\Omega$ is read as a quality signal about the focal firm predicts a uniform sign across peers, because all observers update the same prior. A philanthropy account predicts a concentrated valuation shock at the focal firm with no systematic peer effect. A Bertrand-predation story predicts uniform negative rival response across the full peer set, contradicting the observed null on substitute peers. A network-bootstrap account predicts mixed signs driven by network position rather than by segment exposure, and empirically would not align with a text-based product-market classifier. The asymmetric response is therefore a discriminating prediction, stated as a primary hypothesis rather than as an ex-post heterogeneity result:

**Hypothesis 2 (Asymmetric Peer Valuation):** *The cumulative abnormal return triggered by a strategic open-source release is concentrated on complement peers as a strong negative response and is statistically indistinguishable from zero on substitute peers. The asymmetry is a structural implication of the multi-segment firm's portfolio composition and does not require any assumption about the focal firm's targeting intent.*

Hypothesis 2 is the load-bearing prediction of the framework. It is tested directly in Section 6.3 using the focal firm's peer set partitioned into substitute and complement groups. The empirical operationalization, detailed in Section 6.1, projects the latent exposure pair $(\sigma_r, \gamma_r)$ onto two observable signals: the sign of the peer's pre-event return co-movement with the focal firm and the technological overlap between the peer's patent portfolio and the released repository. Complement peers are identified as peers whose pre-event returns co-move positively with the focal firm's and whose technology overlap with the released repository is high. Substitute peers are identified as peers whose pre-event returns co-move negatively with the focal firm's. This projection loses information relative to a direct measure of segment revenue composition but has the advantage of using only pre-event data, preserving identification.

## A.4 Mechanism Chain in Complement Peers

Hypothesis 2 pins down the equilibrium response of peer market valuations. The mechanism through which the negative response on complement peers propagates into observable firm behavior unfolds across four real-economy margins, each corresponding to one of the four channels in equation (A-3).

A complement peer facing a credible permanent commoditization of its $C$-segment product experiences a structural reduction in pricing power. Revenue scaled by total assets contracts. At the same time, Alexy et al. (2018) observe that the release of an open-source public good by one firm provides a free, high-quality technological input to others, allowing them to reduce the internal expenditure required to produce equivalent capability. Complement peers adopting $\Omega$ as an input substitute cut their cost of goods sold at a rate that, on a scaled basis, can exceed the revenue contraction in the short run. Operating margins therefore expand in the near term even as the peer's competitive position deteriorates.

**Hypothesis 3 (Financial Compression and Short-Run Margin Expansion):** *Complement peers exposed to a focal firm's open-source release exhibit declining sales and declining cost of goods sold, with the cost contraction generating a short-run expansion in operating margins.*

The declining marginal return to proprietary $C$-innovation, together with the continuing availability of $\Omega$ as an evolving baseline maintained by the focal firm, makes internal exploratory research a dominated strategy for the complement peer. Capital reallocates away from exploratory R&D toward exploitation of existing product lines, and the peer's patent portfolio ages as new exploratory filings fall.

**Hypothesis 4 (Innovation Retreat):** *Complement peers reduce internal research and development expenditure and experience aging of their exploratory patent portfolios following a focal firm's open-source release.*

The engineering labor force of a complement peer must maintain compatibility with the evolving open-source standard. In the platform-economics tradition of Gawer and Henderson (2007), standard-setting platforms redirect the focal effort of dependent developers toward the standard itself. In the open-source setting, this redirection is directly measurable as contribution activity on the focal firm's public repositories, separate from proprietary effort on the peer's own codebase.

**Hypothesis 5 (Labor Co-optation):** *Complement peer developers increase active contribution, measured through code commits, pull requests, and issues, to the focal firm's open-source repositories following the release.*

The combined retreat from internal innovation and the reallocation of engineering labor imply that the complement peer's product vocabulary, as captured in its 10-K product description, becomes less dynamic over time. Its product-market fluidity, in the sense of Hoberg et al. (2014), falls.

**Hypothesis 6 (Product-Market Freeze):** *Complement peers experience a decline in product-market fluidity following a focal firm's open-source release.*

Hypotheses 3 through 6 map one-to-one onto the four channels in equation (A-3) and together characterize the complement peer's full adjustment path. The claim is not that all four must appear in every release but that the four move together whenever the release is economically consequential in the sense of the Kogan intensity screen described in Section 4.1. The baseline real-outcome specifications in Section 5 pool substitute and complement peers and therefore estimate an average treatment effect that, under Proposition 2, is driven predominantly by the complement-peer subsample. Section 6.2 decomposes the Sales and COGS responses along the same partition and verifies that the pooled estimates indeed load on the complement pool.

## A.5 Welfare Considerations

The framework delivers signed predictions on producer surplus but leaves the consumer-welfare question open. On the producer side, the focal firm gains by Hypothesis 1, complement peers lose by Hypothesis 2, and substitute peers are approximately unchanged by Hypothesis 2. Summed across the three producer groups, the net effect on producer surplus is negative in every empirical estimate the paper reports, with the complement-peer loss exceeding the focal-firm gain by roughly an order of magnitude in dollar terms. The paper documents this transfer rather than attempting to compute a social-welfare net. On the consumer side, users of segment-$C$ goods obtain a free substitute of quality $V_\Omega$, which is weakly preferred to any priced alternative. Users of segment-$A$ goods benefit from an expanded $C$-market that reduces the effective price of the $A+C$ bundle. Against these gains, the long-run consequence of standards lock-in through channel (iii) can reduce consumer variety by foreclosing entrants whose alternative interfaces would otherwise have emerged. The long-run sign on consumer surplus is therefore ambiguous. Consistent with the empirical stylebook of Seamans and Zhu (2014) and Wen and Zhu (2019), the paper positions its contribution as a documentation of the positive economic fact of the wealth transfer and leaves the welfare question open for downstream analysis.

# B Appendix Tables

**Table IA.1. Chronological List of Top 100 Open Source Releases**

This table provides a chronological record of the top 100 open-source repositories released by focal technology firms between 2015 and 2024. Total Stars and Total Forks reflect the cumulative engagement metrics as of December 2024. Repositories are selected based on their total star counts within each firm's official GitHub organization.

| Release Date | Company | Repository | Total Stars | Total Forks |
|---|---|---|---:|---:|
| 2015-01-02 | Microsoft | Microsoft/TypeScript | 98,724 | 17,110 |
| 2015-01-03 | Adobe | adobe/brackets | 21,495 | 7,447 |
| 2015-01-06 | Alphabet | google/protobuf | 29,385 | 9,284 |
| 2015-01-07 | Alphabet | google/iosched | 20,384 | 7,568 |
| 2015-01-07 | Meta | facebook/immutable-js | 24,331 | 1,867 |
| 2015-01-08 | Alphabet | angular/angular | 117,410 | 37,581 |
| 2015-01-09 | Alphabet | angular/angular.js | 43,059 | 29,601 |
| 2015-01-09 | Netflix | Netflix/Hystrix | 24,237 | 5,388 |
| 2015-01-15 | Alphabet | golang/go | 140,768 | 24,778 |
| 2015-01-16 | Meta | facebook/flow | 22,609 | 2,557 |
| 2015-01-21 | Meta | facebook/folly | 27,143 | 6,346 |
| 2015-01-21 | Meta | facebook/jest | 43,326 | 8,433 |
| 2015-02-02 | Alphabet | google/ExoPlayer | 23,176 | 7,472 |
| 2015-02-04 | Alphabet | google/dagger | 18,953 | 2,762 |
| 2015-02-09 | Meta | facebook/rocksdb | 28,241 | 6,936 |
| 2015-02-11 | Alphabet | google/fonts | 20,706 | 3,894 |
| 2015-02-12 | Alphabet | google/guava | 56,084 | 15,081 |
| 2015-02-22 | Meta | facebook/react | 266,223 | 70,077 |
| 2015-03-26 | Meta | facebook/react-native | 142,423 | 34,803 |
| 2015-03-26 | Meta | facebook/fresco | 20,076 | 5,363 |
| 2015-04-01 | Alphabet | google/flatbuffers | 23,342 | 4,184 |
| 2015-04-13 | Alphabet | google/gson | 26,959 | 6,327 |
| 2015-04-25 | Microsoft | Azure/azure-quickstart-templates | 15,685 | 19,794 |
| 2015-05-13 | Alphabet | google/material-design-icons | 46,021 | 12,892 |

*Continued on next page*

*Table IA1 Continued from previous page*

| Release Date | Company | Repository | Total Stars | Total Forks |
|---|---|---|---:|---:|
| 2015-05-20 | Alphabet | google/styleguide | 41,311 | 17,488 |
| 2015-06-04 | Alphabet | angular/angular-cli | 31,796 | 16,364 |
| 2015-06-11 | Meta | facebook/infer | 16,768 | 2,755 |
| 2015-07-06 | Alphabet | google/material-design-lite | 39,243 | 8,411 |
| 2015-08-10 | Alphabet | kubernetes/kubernetes | 113,262 | 47,065 |
| 2015-08-11 | Meta | facebook/relay | 20,745 | 2,499 |
| 2015-08-17 | Meta | facebook/flux | 16,723 | 5,361 |
| 2015-08-21 | Alphabet | google/googletest | 38,993 | 13,986 |
| 2015-11-04 | Nvidia | NVIDIA/nvidia-docker | 18,562 | 2,626 |
| 2015-11-09 | Alphabet | tensorflow/tensorflow | 222,055 | 113,229 |
| 2015-11-09 | Alphabet | flutter/flutter | 193,431 | 39,867 |
| 2015-11-18 | Microsoft | Microsoft/vscode | 164,552 | 27,778 |
| 2015-12-03 | Apple | apple/swift-evolution | 16,739 | 3,472 |
| 2015-12-03 | Apple | apple/swift | 82,065 | 15,594 |
| 2016-01-04 | Alphabet | angular/material2 | 19,525 | 6,152 |
| 2016-01-25 | Microsoft | Microsoft/CNTK | 17,789 | 5,202 |
| 2016-02-05 | Alphabet | tensorflow/models | 88,698 | 57,029 |
| 2016-02-22 | Meta | facebook/draft-js | 24,194 | 3,527 |
| 2016-04-15 | Alphabet | kubernetes/minikube | 31,350 | 5,918 |
| 2016-05-06 | Alphabet | google/flexbox-layout | 19,971 | 2,202 |
| 2016-06-27 | Alphabet | kubernetes/kops | 16,971 | 5,745 |
| 2016-07-22 | Meta | facebookincubator/create-react-app | 44,224 | 9,725 |
| 2016-08-04 | Meta | facebookresearch/fastText | 28,161 | 5,631 |
| 2016-08-11 | Alphabet | google/leveldb | 39,293 | 9,746 |
| 2016-09-01 | Meta | facebook/zstd | 23,885 | 2,514 |
| 2017-02-21 | Alphabet | google/python-fire | 29,144 | 1,930 |
| 2017-02-28 | Meta | facebookresearch/faiss | 32,476 | 4,116 |
| 2017-04-18 | Alphabet | flutter/plugins | 19,264 | 12,204 |
| 2017-05-09 | Meta | facebookresearch/fairseq | 18,139 | 3,267 |

*Continued on next page*

*Table IA1 Continued from previous page*

| Release Date | Company | Repository | Total Stars | Total Forks |
|---|---|---|---|---|
| 2017-10-06 | Alphabet | kubernetes/ingress-nginx | 17,430 | 9,112 |
| 2018-01-22 | Meta | facebookresearch/Detectron | 29,146 | 6,544 |
| 2018-01-22 | Meta | facebook/create-react-app | 72,792 | 30,901 |
| 2018-02-01 | Oracle | oracle/graal | 20,618 | 2,045 |
| 2018-03-31 | Alphabet | tensorflow/tfjs | 18,519 | 2,379 |
| 2018-06-08 | Alphabet | flutter/samples | 18,703 | 9,245 |
| 2018-08-03 | Alphabet | google/filament | 18,435 | 2,229 |
| 2018-11-18 | Alphabet | google/jax | 30,747 | 3,311 |
| 2019-03-06 | Microsoft | Microsoft/calculator | 33,678 | 5,250 |
| 2019-05-06 | Microsoft | microsoft/vscode | 219,500 | 60,918 |
| 2019-05-06 | Microsoft | microsoft/PowerToys | 122,816 | 8,839 |
| 2019-05-07 | Microsoft | microsoft/TypeScript | 122,588 | 19,046 |
| 2019-05-07 | Microsoft | microsoft/Terminal | 68,718 | 6,328 |
| 2019-05-14 | Microsoft | microsoft/calculator | 34,478 | 9,648 |
| 2019-05-14 | Microsoft | microsoft/vcpkg | 20,043 | 7,528 |
| 2019-05-14 | Microsoft | microsoft/terminal | 145,188 | 17,004 |
| 2019-06-16 | Alphabet | google/mediapipe | 26,152 | 5,477 |
| 2019-06-23 | Meta | facebook/docusaurus | 47,124 | 10,857 |
| 2019-06-24 | Microsoft | microsoft/monaco-editor | 27,804 | 3,050 |
| 2019-09-04 | Alphabet | google/eng-practices | 20,548 | 2,080 |
| 2019-09-18 | Microsoft | microsoft/cascadia-code | 28,595 | 1,120 |
| 2019-09-30 | Microsoft | microsoft/unilm | 20,440 | 2,730 |
| 2019-10-10 | Meta | facebookresearch/detectron2 | 32,410 | 8,868 |
| 2020-01-22 | Microsoft | microsoft/playwright | 69,346 | 4,500 |
| 2020-02-07 | Microsoft | microsoft/DeepSpeed | 35,700 | 4,386 |
| 2020-05-14 | Meta | facebookexperimental/Recoil | 20,714 | 1,486 |
| 2020-05-19 | Microsoft | microsoft/winget-cli | 25,605 | 1,856 |
| 2020-11-10 | Microsoft | microsoft/Web-Dev-For-Beginners | 91,082 | 14,590 |
| 2021-03-03 | Microsoft | microsoft/ML-For-Beginners | 73,726 | 15,665 |

*Continued on next page*

*Table IA1 Continued from previous page*

| Release Date | Company | Repository | Total Stars | Total Forks |
|---|---|---|---|---|
| 2021-03-03 | Microsoft | microsoft/Data-Science-For-Beginners | 28,586 | 6,107 |
| 2021-03-03 | Microsoft | microsoft/AI-For-Beginners | 34,361 | 6,182 |
| 2021-05-05 | Alphabet | google/zx | 34,004 | 1,202 |
| 2022-04-13 | Meta | facebook/lexical | 19,937 | 1,961 |
| 2022-12-21 | Alphabet | google/comprehensive-rust | 27,666 | 1,835 |
| 2023-02-24 | Meta | facebookresearch/llama | 49,019 | 8,188 |
| 2023-03-01 | Microsoft | microsoft/semantic-kernel | 22,163 | 3,501 |
| 2023-03-09 | Microsoft | microsoft/visual-chatgpt | 29,479 | 2,779 |
| 2023-03-31 | Microsoft | microsoft/JARVIS | 23,005 | 2,010 |
| 2023-04-05 | Meta | facebookresearch/segment-anything | 47,319 | 5,795 |
| 2023-06-09 | Meta | facebookresearch/audiocraft | 20,487 | 2,188 |
| 2023-08-18 | Microsoft | microsoft/autogen | 35,473 | 5,401 |
| 2023-10-26 | Microsoft | microsoft/generative-ai-for-beginners | 65,001 | 33,310 |
| 2024-02-27 | Cloudflare | cloudflare/pingora | 22,061 | 1,269 |
| 2024-04-13 | Microsoft | microsoft/MS-DOS | 18,194 | 3,123 |
| 2024-04-18 | Meta | meta-llama/llama3 | 26,665 | 3,108 |
| 2024-07-01 | Microsoft | microsoft/graphrag | 20,517 | 2,045 |
| 2024-12-06 | Microsoft | microsoft/markitdown | 27,821 | 1,092 |

**Table IA.2. Face Validity of TNIC Peer Matching for Major Focal Firms**

This table presents representative examples of peer firms matched to the major tech giants (focal firms) based on the Hoberg-Phillips TNIC product market fluidity data. For each focal firm, this table displays two peers from the High, Mid, and Low competitive threat groups.

| Threat Level | Peer Firm Name | Broad Industry | TNIC Score |
|---|---|---|---|
| **Focal Firm: Microsoft** | | | |
| High | Polycom Inc | Electronic Equipment | 0.1384 |
| | Broadsoft Inc | Software & IT Services | 0.1167 |
| Mid | Yodlee Inc | Software & IT Services | 0.0311 |
| | Ubiquiti Inc | Electronic Equipment | 0.0308 |
| Low | Realpage Inc | Software & IT Services | 0.0002 |
| | Nuance Communications Inc | Software & IT Services | 0.0002 |
| **Focal Firm: Alphabet (Google)** | | | |
| High | Reachlocal Inc | Software & IT Services | 0.0524 |
| | Yelp Inc | Software & IT Services | 0.0302 |
| Mid | Amazon.Com Inc | Online Retail & Web | 0.0161 |
| | Logmein Inc | Software & IT Services | 0.0153 |
| Low | Chegg Inc | Software & IT Services | 0.0035 |
| | Zulily Inc | Online Retail & Web | 0.0009 |
| **Focal Firm: Meta (Facebook)** | | | |
| High | Valero Energy Corp | Other Services | 0.1721 |
| | Spectra Energy Corp | Other Services | 0.1428 |
| Mid | Regency Energy Partners Lp | Other Services | 0.0358 |
| | Northern Tier Energy Lp | Other Services | 0.0349 |
| Low | Freeport-Mcmoran Inc | Other Services | 0.0015 |
| | Seacor Holdings Inc | Other Services | 0.0015 |

**Table IA.2. Face Validity of TNIC Peer Matching for Major Focal Firms (Continued)**

| Threat Level | Peer Firm Name | Broad Industry | TNIC Score |
|---|---|---|---|
| **Focal Firm: Apple** | | | |
| High | Progress Software Corp | Software & IT Services | 0.0443 |
| | Adobe Inc | Software & IT Services | 0.0415 |
| Mid | Astea International Inc | Software & IT Services | 0.0111 |
| | Zagg Inc | Other Services | 0.0107 |
| Low | Autoweb Inc | Software & IT Services | 0.0007 |
| | Tivo Corp | Software & IT Services | 0.0002 |
| **Focal Firm: Nvidia** | | | |
| High | Marvell Technology Inc | Electronic Equipment | 0.1154 |
| | Qualcomm Inc | Electronic Equipment | 0.0818 |
| Mid | Vitesse Semiconductor Corp | Electronic Equipment | 0.0177 |
| | Omnivision Technologies Inc | Electronic Equipment | 0.0177 |
| Low | F5 Inc | Software & IT Services | 0.0005 |
| | Infinera Corp | Electronic Equipment | 0.0002 |
| **Focal Firm: Cloudflare** | | | |
| High | Trinity Industries Inc | Manufacturing | 0.1023 |
| | Martin Marietta Materials | Other Services | 0.0294 |
| Mid | Norfolk Southern Corp | Other Services | 0.0190 |
| | Gatx Corp | Other Services | 0.0090 |
| Low | Canadian Pac Kansas City Ltd | Other Services | 0.0068 |
| | Phillips 66 | Other Services | 0.0021 |

**Table IA.3. Matching Robustness: Full, PSM, CEM, and Entropy Balance**

This table reports Kogan-weighted stacked Difference-in-Differences (DiD) estimates for the three main outcomes—Sales (Panel A), ROA (Panel B), and R&D (Panel C)—under four alternative sample constructions. Column (1) uses the full unmatched sample. Column (2) uses the baseline Propensity Score Matching (PSM) sample. Column (3) applies Coarsened Exact Matching (CEM). Column (4) uses Entropy Balancing (EB). All dependent variables are normalized by total assets. All specifications include Cohort-Firm and Cohort-Time fixed effects, firm controls, and cluster standard errors at the firm level. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans 2015–2024.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Matching Method = | Full | PSM | CEM | EB |
| **Panel A: Dependent Variable = Sales** | | | | |
| Treated × Post | -0.048$^{**}$ | -0.060$^{***}$ | -0.050$^{***}$ | -0.034 |
| | (-2.48) | (-2.69) | (-3.16) | (-1.49) |
| Adjusted $R^2$ | 0.670 | 0.735 | 0.706 | 0.708 |
| Mean of Dep. Var. | 0.723 | 0.728 | 0.723 | 0.723 |
| Observations | 3,898,991 | 974,277 | 3,887,663 | 3,892,496 |
| **Panel B: Dependent Variable = ROA** | | | | |
| Treated × Post | 0.058$^{*}$ | 0.053$^{**}$ | 0.065$^{**}$ | -0.005 |
| | (1.94) | (2.00) | (2.43) | (-0.10) |
| Adjusted $R^2$ | 0.500 | 0.589 | 0.380 | 0.313 |
| Mean of Dep. Var. | -0.095 | -0.064 | -0.089 | -0.095 |
| Observations | 3,898,991 | 974,277 | 3,887,663 | 3,892,496 |
| **Panel C: Dependent Variable = R&D** | | | | |
| Treated × Post | -0.014$^{***}$ | -0.014$^{***}$ | -0.013$^{***}$ | -0.012$^{***}$ |
| | (-3.21) | (-2.77) | (-3.63) | (-3.26) |
| Adjusted $R^2$ | 0.786 | 0.788 | 0.695 | 0.701 |
| Mean of Dep. Var. | 0.093 | 0.107 | 0.092 | 0.093 |
| Observations | 3,900,467 | 974,385 | 3,888,670 | 3,893,505 |

*Note:* All models include Controls, Cohort-Firm FE, Cohort-Time FE, and Firm Clustering.

**Table IA.4. Weighting Robustness: Continuous vs. Rank-Based Kogan Weights**

This table addresses the concern that using the continuous absolute Kogan (2017) innovation value as a weight could be driven by a small number of extreme cohorts. The columns compare the baseline continuous-weight specification to an alternative that replaces the weight with the within-cohort rank (quartile rank, rescaled). The dependent variables are Sales, Return on Assets (ROA), and R&D expenditure, all normalized by total assets. Coefficients are robust across the three main outcomes, confirming that the main result is not driven by extreme-intensity observations. All specifications include Cohort-Firm and Cohort-Time fixed effects, firm-level controls, and cluster standard errors at the firm level. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans 2015–2024.

| Dependent Variable = | Sales | | ROA | | R&D | |
|---|---|---|---|---|---|---|
| Weighting Scheme = | (1) Continuous | (2) Rank-Based | (3) Continuous | (4) Rank-Based | (5) Continuous | (6) Rank-Based |
| Treated × Post | -0.060$^{***}$ | -0.050$^{**}$ | 0.053$^{**}$ | 0.052$^{**}$ | -0.014$^{***}$ | -0.014$^{***}$ |
| | (-2.69) | (-2.48) | (2.00) | (2.07) | (-2.77) | (-2.99) |
| Adjusted $R^2$ | 0.735 | 0.739 | 0.589 | 0.476 | 0.788 | 0.745 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustering | Firm | Firm | Firm | Firm | Firm | Firm |
| Mean of Dep. Var. | 0.728 | 0.728 | -0.064 | -0.064 | 0.107 | 0.107 |
| Observations | 974,277 | 974,277 | 974,277 | 974,277 | 974,385 | 974,385 |

**Table IA.5.**

**Illustrative Examples of Substitute vs. Complement Peers (CPC × Corr Classification)**

This table illustrates how the two-dimensional classifier assigns peer firms into *Substitute* vs. *Complement* cohorts for three well-known open-source releases by focal technology firms. A peer is labeled a *Substitute* if its CPC patent cosine similarity with the focal repository exceeds the cohort median **and** its pre-event $[t-250, t-30]$ industry-adjusted return correlation with the focal firm is *negative*; a peer is labeled a *Complement* if CPC similarity is above the cohort median **and** the pre-event correlation is *positive*. Peer examples are drawn from the top-CPC-similarity pool of each cohort; focal firms and firms appearing in multiple panels have been excluded to showcase distinct narrative cases. *CPC Sim.* is cosine similarity (0–1) between the peer's patent portfolio and the focal repository. *Corr* is the peer's pre-event $[t-250, t-30]$ industry-adjusted return correlation with the focal firm. Peer announcement-window CAR — the object the classification is designed to predict — is reported separately in Slide 5.3 / Section 5. Sample period: 2015–2024.

| **Substitute Peers (CPC > med. & Corr < 0)** | | | **Complement Peers (CPC > med. & Corr > 0)** | | |
|---|---|---|---|---|---|
| Peer Firm | CPC Sim. | Corr | Peer Firm | CPC Sim. | Corr |
| **Panel A: microsoft/vscode** | | | (Focal: Microsoft Corp.) — *free developer IDE* | | |
| Splunk Inc. | 0.98 | –0.02 | Oracle Corp. | 0.98 | +0.25 |
| Appian Corp. | 0.96 | –0.09 | MongoDB Inc. | 0.94 | +0.11 |
| Carbon Black Inc. | 0.86 | –0.22 | PTC Inc. | 0.92 | +0.25 |
| **Panel B: facebook/react** | | | (Focal: Meta Platforms Inc.) — *free UI framework* | | |
| LivePerson Inc. | 0.77 | –0.06 | Twitter Inc. | 0.87 | +0.24 |
| Zendesk Inc. | 0.73 | –0.12 | LinkedIn Corp. | 0.85 | +0.22 |
| TigerLogic Corp. | 0.64 | –0.01 | Yelp Inc. | 0.78 | +0.22 |
| **Panel C: microsoft/deepspeed** | | | (Focal: Microsoft Corp.) — *free LLM-training framework* | | |
| Pivotal Software Inc. | 0.96 | –0.09 | VMware Inc. | 0.96 | +0.10 |
| Cloudera Inc. | 0.95 | –0.13 | Smartsheet Inc. | 0.95 | +0.07 |
| Alteryx Inc. | 0.90 | –0.02 | Teradata Corp. | 0.91 | +0.04 |

**Table IA.6. Robustness: Substitute vs. Complement Peers under 2D Classification**

This table replicates the main Substitute-vs-Complement analysis using the 2D classification, which intersects CPC technology overlap with industry-adjusted stock-return correlation and drops off-diagonal peers. Pure Substitute peers exhibit high CPC overlap *and* negative return co-movement; Pure Complement peers exhibit high CPC overlap *and* positive return co-movement. The dependent variables Sales and COGS are normalized by total assets. All specifications include Cohort-Firm and Cohort-Time fixed effects and firm-level controls, and observations are weighted by the absolute value of the cohort-level Kogan (2017) innovation value. Standard errors are clustered at the firm level. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample period spans 2015–2024.

| Dependent Variable = | Sales | | COGS | |
|---|---|---|---|---|
| Peer Firm Type = | (1) Substitute | (2) Complement | (3) Substitute | (4) Complement |
| Treated × Post | -0.0545 | -0.1211$^{***}$ | -0.0466$^{*}$ | -0.0765$^{***}$ |
| | (-1.60) | (-3.69) | (-1.82) | (-2.86) |
| Adjusted $R^2$ | 0.630 | 0.681 | 0.627 | 0.701 |
| Controls | Yes | Yes | Yes | Yes |
| Cohort-Firm FE | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes |
| Clustering | Firm | Firm | Firm | Firm |
| Weight | Yes | Yes | Yes | Yes |
| Mean of Dep. Var. | 0.681 | 0.700 | 0.315 | 0.338 |
| Observations | 422,642 | 448,366 | 422,642 | 448,366 |

**Table IA.7. Peer GitHub Activity Response: Unweighted Robustness**

This table reports the unweighted versions of the Labor co-optation specification in main Table 5. Panel A uses the full Kogan-eligible cohort universe; Panel B restricts to the Top 100 BigQuery-linked sub-sample. Both panels exclude the Kogan-intensity weights so that each focal release contributes equally to the estimate. Outcomes are identical to those in Table 5: per-developer activity (total, active, and passive contributions) and the size of the active developer base. The unweighted estimates are larger in magnitude and tighter in significance than the Kogan-weighted main results, indicating that the average effect across all releases is at least as strong as the average across the most economically influential releases. The passive-engagement effect, which is statistically indistinguishable from zero in the weighted main specification, becomes positive and significant when each cohort enters the estimate equally. All specifications include firm-level controls, cohort-firm and cohort-time fixed effects, and firm-clustered standard errors. $t$-statistics are reported in parentheses. $^{***}$, $^{**}$, and $^{*}$ indicate statistical significance at the 1%, 5%, and 10% levels.

| Dependent Variable = | Activity per Developer | | | Developer Base |
|---|---|---|---|---|
| | (1) Any Work per Dev. | (2) Active Contrib. | (3) Passive Contrib. | (4) Ln(Working Devs + 1) |
| *Panel A: Full sample* | | | | |
| Treated × Post | 1.593$^{***}$ | 1.502$^{**}$ | 0.091$^{***}$ | 0.245$^{***}$ |
| | (2.62) | (2.47) | (3.47) | (8.02) |
| Adjusted $R^2$ | 0.708 | 0.705 | 0.404 | 0.816 |
| Mean of Dep. Var. | 3.122 | 2.801 | 0.321 | 0.480 |
| Observations | 974,385 | 974,385 | 974,385 | 974,385 |
| *Panel B: Top 100 sub-sample* | | | | |
| Treated × Post | 0.419$^{***}$ | 0.378$^{***}$ | 0.042$^{*}$ | 0.171$^{***}$ |
| | (3.58) | (3.47) | (1.65) | (2.93) |
| Adjusted $R^2$ | 0.116 | 0.075 | 0.339 | 0.729 |
| Mean of Dep. Var. | 1.071 | 0.756 | 0.315 | 0.584 |
| Observations | 49,760 | 49,760 | 49,760 | 49,760 |
| Controls | Yes | Yes | Yes | Yes |
| Cohort-Firm FE | Yes | Yes | Yes | Yes |
| Cohort-Time FE | Yes | Yes | Yes | Yes |
| Clustering | Firm | Firm | Firm | Firm |
| Weight | No | No | No | No |

**Table IA.8. Focal Firm Demand Response: Falsification Tests on the Within-Firm Specification**

This table reports three placebo / falsification tests on the within-firm core-vs-non-core static DiD coefficient reported in Table 10 column (1). P1 randomly shifts each focal firm's first-release year by an integer drawn from $[-5,+5]\setminus\{0\}$ and re-estimates the $\textit{IsCore} \times \textit{Post}$ coefficient; 500 perturbations are drawn. P2 permutes the first-release year across focal firms (so that Microsoft's event year is assigned to Adobe, and so on); valid (non-identity) permutations only, target 500 draws. P3 randomly reshuffles which segments are flagged as core within each focal firm while preserving the firm-level count of core segments; 500 perturbations. The reported two-sided $p$-value is the share of placebo coefficients with magnitude $\geq$ the actual point estimate. P2 returns zero valid permutations because all eight focal firms first released in calendar year 2015 (an artifact of the GitHub-adoption ramp documented in Figure 1), which makes the cross-firm date swap degenerate; this clustering is itself diagnostic of why the within-firm column (1) is underpowered and motivates the triple-DiD and continuous-dose specifications in columns (2)–(4) of Table 10. The corresponding null distributions are plotted in Figure IA3.

| Placebo design | Null mean | Null SD | Actual coef | Two-sided $p$ |
|---|---|---|---|---|
| P1: random ±5-year shift | +0.282 | 0.227 | +0.193 | 0.700 |
| P2: cross-firm date swap | n/a | n/a | +0.193 | n/a |
| P3: wrong-segment within-firm shuffle | +0.004 | 0.215 | +0.193 | 0.400 |

## B.1 Selection and Attenuation: Saturation Curve and Trimming

The headline stacked difference-in-differences specification weights all 1,671 release cohorts by their ex-ante Kogan intensity $|\kappa_r|$. Two reviewer-relevant concerns merit a direct check. First, the set of included cohorts is fixed by mechanical rules, namely GitHub Archive coverage, a Compustat focal match, and the availability of a one-to-one propensity-score-matched control; a reviewer may worry that these rules implicitly select on outcome-relevant characteristics, particularly relative to a cutoff defined by cumulative repository popularity. Second, low-$|\kappa_r|$ releases could add noise that attenuates the weighted estimate. I address both with a saturation curve and a trimming-from-bottom exercise, reported in Table IA.9 and visualized in Figure IA2.

*Saturation curve.* Panel A of Table IA.9 ranks all cohorts by cohort-level $|\kappa_r|$ and re-estimates the headline $T \times Post$ coefficient on the Top-$K$ subset for $K \in \{100, 250, 500, 1000, 1671\}$. The diagnostic is the following. If low-$|\kappa_r|$ releases were noise, the unweighted curve should drift toward zero as $K$ grows while the weighted curve stays put; if instead those releases carry attenuated but directionally consistent signal, both curves should converge as $K$ approaches the full sample. The data show the latter. For Sales/Assets the $|\kappa_r|$-weighted estimate is $-0.065$ at $K = 100$ and $-0.060$ at the full sample, a movement well inside one standard error, whereas the unweighted estimate drifts from $-0.063$ to $-0.046$, an attenuation of roughly 27%. R&D/Assets and ROA behave the same way: the weighted curve is approximately constant in $K$, and the unweighted curve converges to it from above in absolute value as low-$|\kappa_r|$ releases are added. Two implications follow. The substantive effect is not driven by the cumulative-popularity Top 100 cutoff, since restricting to the ex-ante highest-$|\kappa_r|$ Top 100 alone yields nearly the same coefficient as the full set of cohorts; and because the weighted estimate is stable across $K$, the full-sample weighted estimate is not diluted by the inclusion of low-intensity releases.

*Trimming from the bottom.* Panel B drops the bottom $q$% of cohorts by $|\kappa_r|$ and re-estimates the same specification for $q \in \{0, 25, 50, 75\}$. Under attenuation the coefficient should grow in absolute value as $q$ rises. The weighted Sales coefficient is essentially flat across all four trims ($-0.060$, $-0.060$, $-0.062$, $-0.065$), while the unweighted Sales coefficient rises monotonically from $-0.046$ to $-0.062$, exactly the attenuation pattern that motivates the weighting choice. The two estimators meet at $q = 75$, where only the top $|\kappa_r|$ quartile of cohorts remains; the R&D and ROA panels show the same behavior. Low-$|\kappa_r|$ releases therefore contribute proportionally smaller, directionally consistent signal rather than dilutive noise, which is precisely the inverse-variance logic that justifies weighting by $|\kappa_r|$. Consistent with this reading, the per-quartile dose-response estimates of Table IA.4 retain a substantial share of the headline magnitude in the bottom $|\kappa_r|$ quartile, roughly 60% for Sales/Assets, 71% for R&D/Assets, and 55% for ROA.

## Table IA.9. Selection and Attenuation: Saturation Curve and Trimming-from-Bottom

Panel A re-estimates the headline weighted stacked DiD $T \times Post$ coefficient on Top-$K$ subsets of release cohorts ranked by ex-ante Kogan intensity $|\kappa_r|$ (descending), for $K$ from 100 up to the full sample of 1,671 cohorts (the rightmost block is labeled Top-1700 for readability). Panel B reports the same coefficient with the bottom $q\%$ of cohorts by $|\kappa_r|$ dropped. "Wgt" is the $|\kappa_r|$-weighted specification; "Unw" is unweighted. All specifications include Cohort-Firm and Cohort-Time fixed effects and firm-level controls; standard errors are clustered by firm and $t$-statistics are reported in parentheses. The sample is the full stacked DiD panel. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

*Panel A: Saturation curve — Top-K cohorts by $|\kappa|$*

| K | Top-100 | | Top-250 | | Top-500 | | Top-1000 | | Top-1700 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Wgt | Unw | Wgt | Unw | Wgt | Unw | Wgt | Unw | Wgt | Unw |
| Sales/Assets | −0.0649** | −0.0633** | −0.0667*** | −0.0657** | −0.0642*** | −0.0608*** | −0.0606*** | −0.0504** | −0.0597*** | −0.0463** |
| | (-2.56) | (-2.49) | (-2.59) | (-2.52) | (-2.69) | (-2.71) | (-2.69) | (-2.39) | (-2.69) | (-2.39) |
| R&D/Assets | −0.0089 | −0.0092 | −0.0129** | −0.0145** | −0.0139** | −0.0157*** | −0.0141*** | −0.0150*** | −0.0138*** | −0.0125*** |
| | (-1.45) | (-1.53) | (-2.19) | (-2.47) | (-2.56) | (-3.07) | (-2.76) | (-3.04) | (-2.77) | (-2.78) |
| ROA | 0.0378 | 0.0427 | 0.0525* | 0.0600* | 0.0544* | 0.0603** | 0.0537** | 0.0549** | 0.0528** | 0.0474** |
| | ( 1.36) | ( 1.48) | ( 1.73) | ( 1.86) | ( 1.91) | ( 2.17) | ( 1.98) | ( 2.01) | ( 2.00) | ( 2.02) |
| Cohorts (events) | 100 | 100 | 250 | 250 | 500 | 500 | 1000 | 1000 | 1671 | 1671 |
| Obs. | 69,055 | 69,055 | 179,122 | 179,122 | 336,798 | 336,798 | 638,158 | 638,158 | 974,277 | 974,277 |

*Panel B: Trimming-from-bottom — drop bottom q% of cohorts by $|\kappa|$*

| Bottom dropped | $q = 0\%$ | | $q = 25\%$ | | $q = 50\%$ | | $q = 75\%$ | |
|---|---|---|---|---|---|---|---|---|
| | Wgt | Unw | Wgt | Unw | Wgt | Unw | Wgt | Unw |
| Sales/Assets | −0.0597*** | −0.0463** | −0.0600*** | −0.0487** | −0.0616*** | −0.0536** | −0.0650*** | −0.0623*** |
| | (-2.69) | (-2.39) | (-2.69) | (-2.37) | (-2.70) | (-2.50) | (-2.68) | (-2.68) |
| R&D/Assets | −0.0138*** | −0.0125*** | −0.0139*** | −0.0136*** | −0.0142*** | −0.0156*** | −0.0139** | −0.0160*** |
| | (-2.77) | (-2.78) | (-2.76) | (-2.85) | (-2.74) | (-3.14) | (-2.51) | (-3.01) |
| ROA | 0.0528** | 0.0474** | 0.0532** | 0.0519** | 0.0538** | 0.0560** | 0.0539* | 0.0601** |
| | ( 2.00) | ( 2.02) | ( 2.00) | ( 2.02) | ( 1.96) | ( 2.04) | ( 1.86) | ( 2.06) |
| Cohorts (events) | 1671 | 1671 | 1253 | 1253 | 835 | 835 | 417 | 417 |
| Obs. | 974,277 | 974,277 | 762,804 | 762,804 | 547,253 | 547,253 | 286,568 | 286,568 |

Wgt = Kogan $|\kappa|$-weighted; Unw = unweighted. $t$-stats in parentheses, clustered by gvkey.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## B.2 Full-Sample VW CAR Robustness

The main-text market-reaction analysis in Section 6.3 restricts attention to the top 100 strategically influential open-source releases because daily FF4 factor loadings for every peer firm are computationally costly to estimate at scale. This appendix verifies that the qualitative pattern, a positive focal CAR, a statistically insignificant substitute CAR, and a significantly negative complement CAR, survives when the sample is expanded to include every Kogan-eligible release with return coverage. Panel A of Appendix Table IA.10 reports the value-weighted and equal-weighted $CAR_{[-1,+1]}$ estimates for focal firms, substitute peers, and complement peers in the full sample. The focal, substitute, and complement dollar aggregates likewise preserve the direction and magnitude of the targeted wealth transfer documented in the main text. Appendix Figure IA.1 shows the full-sample cumulative average abnormal return trajectory over the extended $[-10,+50]$ window as a visual complement.

**Table IA.10. Full-Sample Robustness: EW vs. VW $CAR_{[-1,+1]}$ by Peer Type**

This table reports equal-weighted (EW) and value-weighted (VW) cumulative abnormal returns over the $[-1,+1]$ event window, along with aggregate dollar-value impact, for the full Kogan-eligible sample (i.e., every release for which peer-firm return data are available, without the top-100 restriction imposed in the main text). Abnormal returns are estimated from the FF4 model. Firms are weighted by pre-event market capitalization in the VW specification. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

| Group | N | EW CAR% | EW *t* | VW CAR% | VW *t* | Agg. $B |
|---|---|---|---|---|---|---|
| Focal Firms | 1,707 | +0.22 | (+2.36)** | +0.17 | (+3.85)*** | +1,905 |
| Substitute Peers | 31,041 | +0.08 | (+1.54) | +0.06 | (+1.42) | +1,055 |
| Complement Peers | 17,402 | −0.13 | (−2.88)*** | −0.13 | (−3.12)*** | −143 |

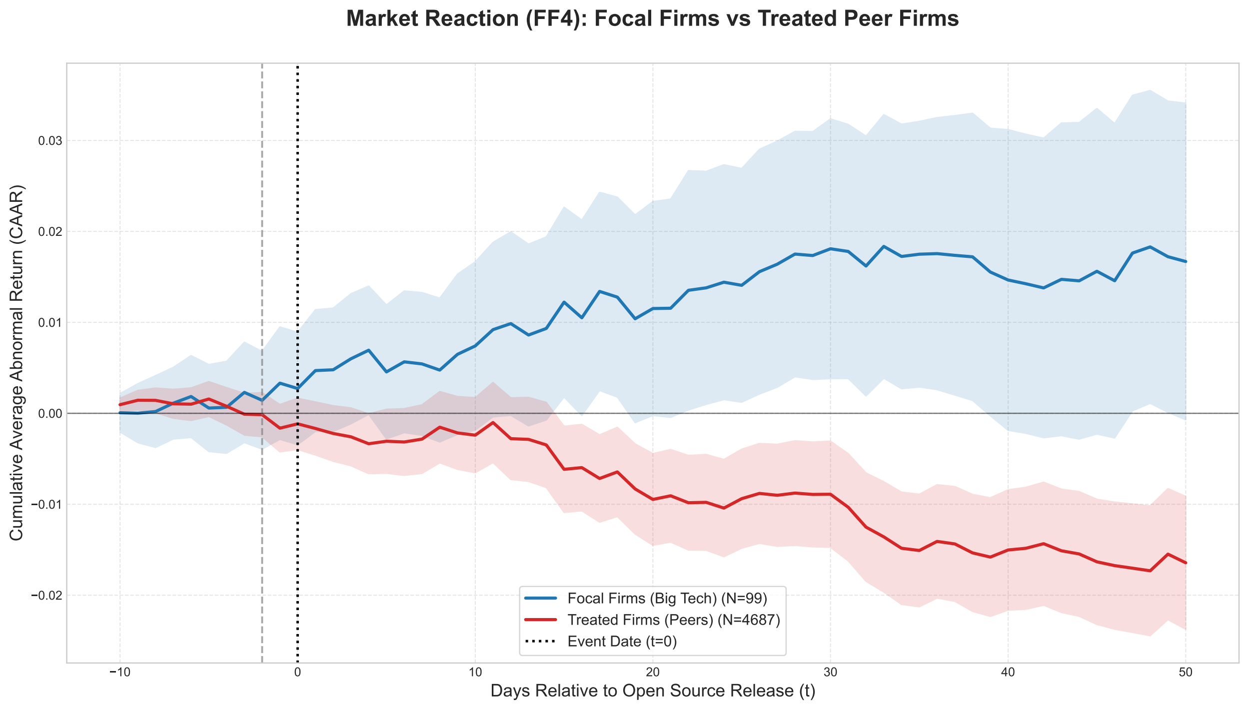



**Figure IA.1. Full-Sample CAAR: Focal vs. Treated Peers over** $[-10, +50]$
Cumulative average abnormal returns (CAAR) for focal tech giants and the pooled treated peer group over the trading-day window $[-10, +50]$ relative to the release day (Day 0), estimated from the FF4 model with 150-day estimation windows ending 30 days prior to the event. Shaded bands represent 90% confidence intervals. Unlike the main-text Figure 10, this version uses the full sample and pools substitute and complement peers; the pooled peer trajectory lies between the two individual partitions documented in the main text.

# C Appendix Figures

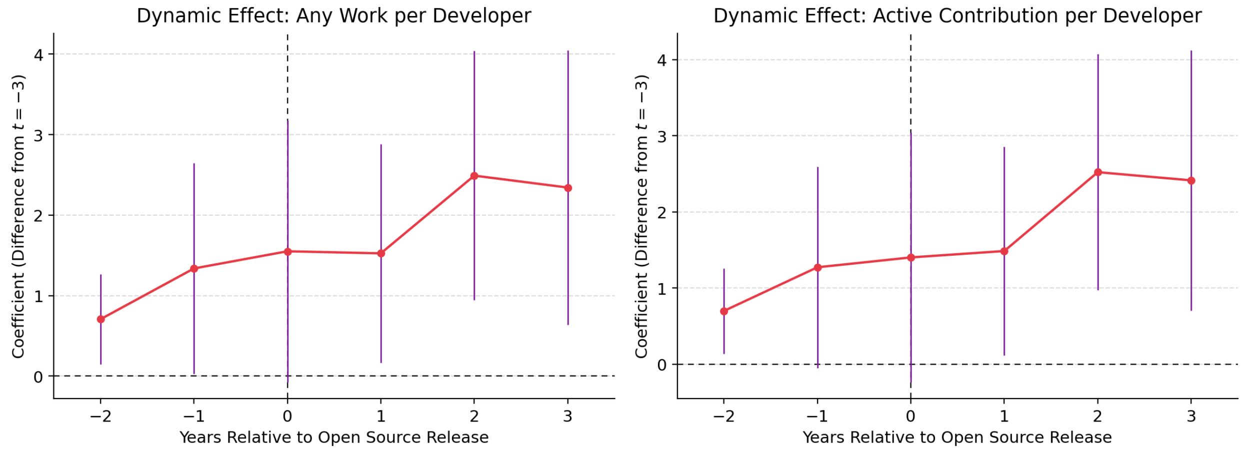



**Figure IA1. Dynamic Effect on Peer Firm GitHub Activity per Developer (Full Sample)**
This figure plots the dynamic treatment effect of focal firms' open-source releases on peer firms' per-developer GitHub activity using the full Kogan-eligible sample of approximately 1,700 cohorts, matching Panel A of Table 5. Panel (A) shows any contribution per working developer; Panel (B) shows active contributions per working developer. The specification is identical to main-text Figure 5 except for the broader sample. As discussed in Section 5.3, the broader sample exhibits a positive pre-treatment trajectory consistent with the secular GitHub-adoption ramp affecting most peer firms during 2014–2017; the pooled DiD estimate in Table 5 Panel A controls for this trend through the cohort $\times$ year fixed effects. The cleaner identification on the Top 100 BigQuery-linked sub-sample is reported in main-text Figure 5, where the pre-treatment coefficients are statistically indistinguishable from zero. Coefficients are plotted relative to $t = -3$ with 95% confidence intervals.

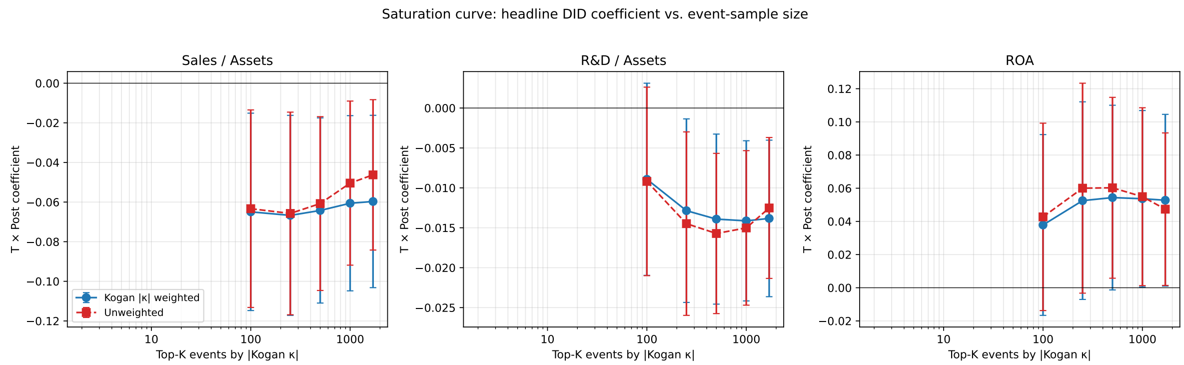



**Figure IA2. Saturation Curve: Headline DiD Coefficient vs. Event-Sample Size**

Each panel plots the weighted stacked-DiD $T \times Post$ coefficient with its 95% confidence interval against the Top-$K$ subset of release cohorts ranked by ex-ante Kogan intensity $|\kappa_r|$, for the three main outcomes (Sales/Assets, R&D/Assets, ROA). The solid line is the $|\kappa_r|$-weighted estimate; the dashed line is the unweighted estimate. The two curves converging as $K$ approaches the full sample of approximately 1,700 cohorts indicates that low-$|\kappa_r|$ releases contribute attenuated but directionally consistent signal rather than diluting the estimate. Companion to Table IA.9; the sample is the full stacked DiD panel.

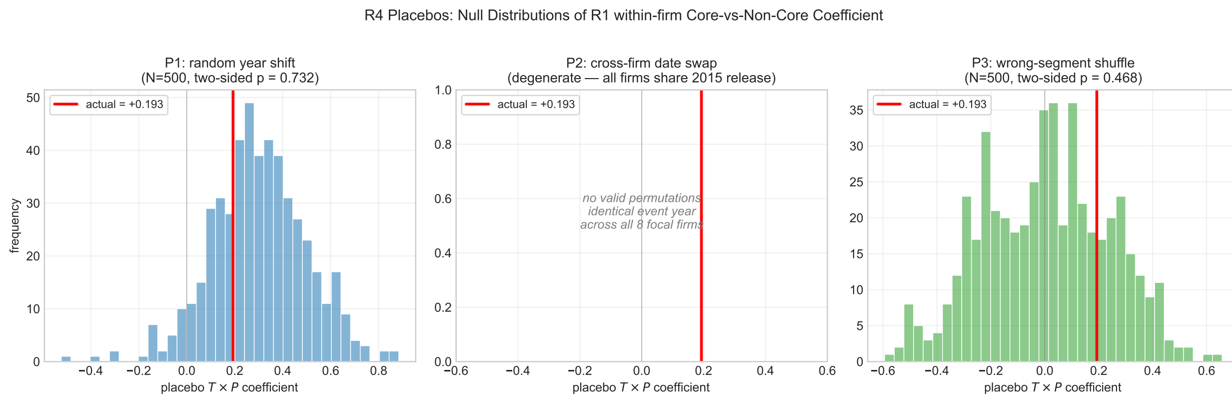



**Figure IA3. Focal Firm Demand Response: Placebo Null Distributions**

Each panel plots the histogram of $IsCore \times Post$ coefficients across 500 perturbations of the auxiliary within-firm core-vs-non-core specification visualized in Figure 12 panel (a). P1 (left) randomly shifts each focal firm's first-release year by an integer in $[-5, +5] \setminus \{0\}$. P2 (centre) permutes first-release years across focal firms; it returns zero valid permutations because all eight focal firms first released in calendar year 2015, an artifact of the GitHub-adoption ramp. P3 (right) randomly reshuffles which segments are flagged as core within each firm while preserving the firm-level count of core segments. The red vertical line marks the actual point estimate ($+0.193$). The P1 null is centred to the right of the actual estimate, reflecting that random year shifts in this clustered-treatment setting mechanically pick up the post-2015 sectoral trend; together with P2's degeneracy, this is exactly why the within-firm binary specification is underpowered on its own. Identification of the focal demand response in the main paper therefore comes from the continuous cumulative-dose specifications in Table 10, where the post-2014 treatment-period specifications reach $t$-statistics above $+4.7$.