The Paradox of Openness: Exposure vs. Efficiency of APIs

Seth G. Benzell*
Guillermo Lagarda†
Jonathan Hersh‡
Marshall Van Alstyne§

August 3, 2019

ABSTRACT

APIs are the building blocks of digital platforms, yet there is little quantitative evidence on their use. Do API adopting firms do better? Do such firms change their operating procedures? Using proprietary data from a major API tools provider, we explore the impact of API use on firm value and operations. We find evidence that API use increases market capitalization and lowers R&D expenditures. We then document an important downside. API adoption increases the risk of data breaches, a risk that rises when APIs are more open or place less emphasis on security. Firms reduce API data flows in the month before a hack announcement, consistent with a conscious attempt to limit breach scope. In the same period, however, the variance of API data flows increases, consistent with heterogeneity in firms' ability to detect and shut down unauthorized data access. Our findings highlight a fundamental paradox of openness: It increases upside value and downside risk at the same time. We document that firms respond to these trade-offs in logical ways and conclude that the benefits of opening APIs exceed the risks for firms situated to adopt a platform strategy.

Keywords: Platforms, APIs, Information Security, Technology Strategy, Market Capitalization

*MIT Initiative on the Digital Economy. Email: sbenzell@mit.edu.
†Inter-American Development Bank. Email: glagarda@idb.org.
‡Chapman University, Argyros School of Business. Email: hersh@chapman.edu.
§Boston University, Questrom School of Business. Email: mva@bu.edu.

I. Introduction

In the information age, the value of a firm rests fundamentally on how it stores, shares and processes information.[1] Digital infrastructure is central to a firm's success. For platform businesses, which rely on creating an ecosystem of interactions and capturing a share of the resulting surplus, this truism holds especially strongly. Such systems harness third party resources they do not own (Parker et al., 2017), enable collaboration between actors and integration among resources (Baldwin and Clark, 2006), and feature a modular architecture of remixable resources (Lusch and Nambisan, 2015). A robust and generative digital system rarely emerges from firms that simply use technology to pave the "digital cow paths" of prior practice (Tilson et al., 2010; Henfridsson and Bygstad, 2013). In this paper, we quantitatively investigate the firm-level consequences of a new practice spreading among firms' digital infrastructures. Specifically, we investigate the performance and operations of firms that implement application programming interfaces (APIs).

As a foundation for a digital infrastructure, APIs offer the dual virtues of practical modular design and precise metering of access. Modular architecture allows designers to independently create, subdivide, modify, and remove components without affecting other parts of a larger system (Baldwin and Clark, 2006). In effect, one can "virtualize" business processes in the manner of virtualizing computer resources (Iyer and Henderson, 2010). This also facilitates partitioning of decision rights (Tiwana et al., 2010). Modularity combines the advantages of standardization typically associated with high volume processes together with the advantages of customization typically associated with bespoke processes (Baldwin and Clark, 2000). APIs also enable precise metering of access permissions to these key resources. Metered access permissions ensure that anyone and anything that consumes system resources adheres to technical and economic policies designed to ensure system health (Jacobson et al., 2011). As architecture, APIs provide scalable infrastructure for building platforms. As regulators, APIs partition decision rights, augmenting the API controller's ability to govern behavior. In fulfilling these

[1]The value of US corporate intangible assets increased from near zero in the early 1990s to more than $6.6 Trillion in 2016 (Benzell and Brynjolfsson, 2019) – calculated as US corporate equity and liabilities less financial assets from Federal Reserve series Z.1 and less fixed capital from BEA table 4.1.

roles – architecture and governance – APIs serve as the foundation for digital platforms (Parker et al., 2016).

Rising interest in APIs has gone hand in hand with the rising dominance of platform firms in the economy. In 2017, six of the top ten firms by market value were platforms.[2] APIs simplify the writing and operation of programs that communicate with online services and shared databases. They are essential for powering such systems as Google's documents and maps, Amazon's voice and web services, Apple's online market, and Facebook's authentication services. They mediate economic transactions. Their value is not only determined by the actions of their creators but also by the habits of their users and the strategic choices of third parties who connect systems and reuse components in unanticipated ways.

As a form of digital infrastructure, the usefulness of APIs depends on how well they balance relevant considerations. An API is a kind of aperture that selects a level of information to diffuse in and out. Too wide an aperture and the firm may give away its competitive advantage. Too narrow or difficult to access and outsiders will struggle to engage an aperture meaningfully.

Prior literature has identified three specific tradeoffs. The first is a balance between adaptability, which improves fit, and stability, which promotes efficiency. Indeed, Baldwin and Woodard (2009) define a platform as a modular complex system in which the core remains stable and the complements evolve. The need to create a stable system that can also adapt is the "Paradox of Change."

The second tradeoff balances opening, in order to encourage outside innovation, against closing in order to monetize. Put differently, too much control drives partners away, but too little control limits value capture. This is the "Paradox of Control" (Tilson et al., 2010; Zeng, 2015).

A recent wave of data breaches highlights a third tradeoff – between an interest in enabling third party innovations and an interest in thwarting third party hacks. Opening APIs can have both effects. The tradeoff depends in part on the relative mix of benevolent and malicious

[2]As measured by the presence or absence of external developers. These firms are Apple (1), Google (2), Microsoft (3), Amazon (4), Facebook (5), and Alibaba (8). Source: `http://dogsofthedow.com/largest-companies-by-market-cap.htm` accessed June 20, 2017. Confirmed June 19, 2019. Source: `https://en.wikipedia.org/wiki/List_of_public_corporations_by_market_capitalization#2019`

outsiders, which is hidden information. Ransbotham (2016) has shown this "Paradox of Exposure" to be present in the context of open-source software.

In this paper, we investigate how firms use APIs and the consequences of using them on firm performance. Despite significant research on the economics of platforms and other types of IT capital, there is little research on APIs themselves. To our knowledge, this is the first paper to statistically investigate the role of APIs in either market value or data breaches.

Our reading of the literature leads us to four testable hypotheses. The first is that API using firms should experience faster growth in market value. Second is that firms get better at managing the tradeoffs inherent in API use over time. The third is that open APIs allow firms to substitute away from internal R&D as third parties create value. The fourth is that APIs, when improperly managed, contribute to firm vulnerability. We test these hypotheses on the event of a firm making its first API call, while controlling for as many alternate theories as we have been able to identify.

Our data set combines proprietary API-level data from a prominent API management company, firm-level financial information from Compustat, and a firm-level data-breach panel from the *Privacy Rights Clearinghouse*. To evaluate our hypothesis, we exploit the heterogeneous timing of API adoption by different firms. This allows us to employ a difference-in-difference research design that controls for time-invariant firm-specific heterogeneity (through firm fixed effects) as well as macroeconomic fluctuations that might correlate with API adoption timing (through time fixed effects). We focus on a leads-and-lags model, as in (Burtch et al., 2018), which allows us to measure changes in the impact of the treatment over time. Using this research design we estimate the change in hack risk and firm financial outcomes due to API adoption, as well as the change in API data flows due to a data breach.

Despite our unique data and careful research design, evaluating our hypotheses still faces challenges. One major limitation is that we only have API usage data for roughly ten dozen firms (123). Another is the fact that API adoption is endogenous, with no obvious available instrument. The consequences of these identification challenges are discussed in more detail throughout the paper, and we address the challenges we can through the use of multiple robustness checks. Despite these limitations, we find significant support for each of our four

motivating hypotheses. Although our identification is less than ideal, the importance, urgency and novelty of our questions make an honest evaluation of the available data essential.

Our paper begins by reviewing the history and literature on APIs. It proceeds to describe and analyze our data. We conclude with a discussion of the tradeoffs involved in the use of APIs, and find that their adoption is well warranted for properly managed firms.

II. Background and Hypotheses

An API is a set of routines, protocols, and tools that standardizes building software applications compatible with an associated program or database. APIs are code. They are also contracts (Jacobson et al., 2011). They govern the type and format of calls or communications that any application can make of another associated program. The answering program is agnostic about the source of the call, yet can require access permission, and the calling program need not know anything about the internal workings of the answering program.

In the same way that user interfaces on operating systems make personal computers easier to use, interfaces on applications make programs easier to use by machines as well as people.[3] Firms use APIs to offer network access to distributed data and services. Such services logically represent a business activity, produce a specific result, are self-contained, and are readily recomposable.[4]

Many web-pioneers featured APIs as core to their business. Salesforce.com included them in their 2000 launch of the world's first software-as-a-service product. Likewise, eBay launched a developer program in 2000 to a select group of partners, encouraging them to create services that drew information from eBay's API. Having created one of the first popular open APIs, eBay's decision led to a virtuous cycle of better tools, higher visibility, and more customers.[5]

It is not clear when the first API was created, but they clearly predate the Internet. Google's ngram tool lists usage of the phrase 'application programming interface' as early as 1961. As

[3]Source: David Berlind, Chief Editor Programmable Web. `http://www.crn.com/news/networking/300084244/att-adds-apis-to-help-partners-move-more-mobility-products-services.htm`

[4]Representative API data types include documents, images, video, geolocation, news feeds, etc.. Representative API functions include ID verification, payment, notification, visualization, language translation, mapping, etc.. `https://en.wikipedia.org/wiki/Service-oriented_architecture`

[5]Op. cit. Berlind

shown in Figure 1, API use has grown exponentially since 2005.

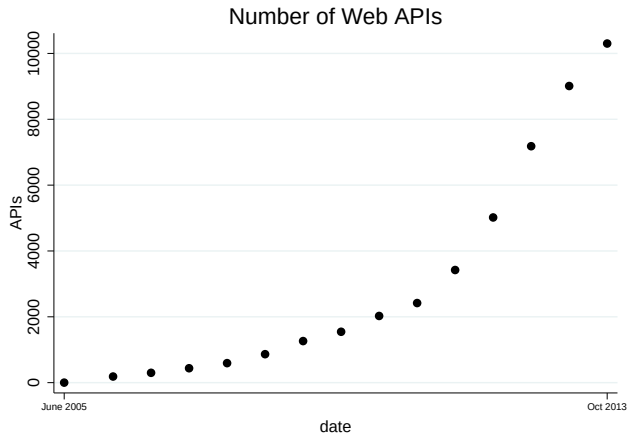


Figure 1: Number of Web APIs over time.
Source: Authors' figure using Programmable Web (2016) data.

There are important differences between APIs and earlier outsourcing, back-office, and front-office technologies such as electronic data interchange (EDI), enterprise resource planning (ERP), and customer relationship management (CRM). First, most prior use cases were for internal use or for use with *known* partners only. By contrast, APIs specifically emphasize "permissionless" innovation by partners the firm never met, who generate projects the firm never conceived (Thierer, 2016; Chesbrough and Van Alstyne, 2015; Parker et al., 2017). Salient illustrations include the numerous apps sold by Apple, Amazon, and Google but that had nothing to do with these platforms themselves.Second, APIs are more than technical plumbing designed to decrease transaction costs or increase efficiency. They *enable* markets. The consequence is not merely a shift from hierarchies to markets or a shift in the 'make-vs-buy' decision (Malone et al., 1987). Instead of entering the market as a more efficient buyer, the focal firm *becomes* a market, an orchestrator of other firms' transactions. Technological market making then has strategic implication as a business sensing function beyond that normally attributed to ERP, CRM and EDI. Orchestrating a market gives the platform visibility into the data passing through its systems, which provides insights even into competitors' activities, margins, and opportunities (Khan, 2017). This yields an information asymmetry that favors the platform sponsor at the expense of the platform partner (Zhu and Liu, 2018). Advantage born of this asymmetry contributes to antitrust scrutiny of platforms in the EU and the US

(Schulze, 2019).

Table I and Figure 2 suggest that use of APIs could influence firm performance. Although insufficient for causal inference, visual inspection of the divergent trends suggests that API usage warrants attention. The special success of Amazon, the one retailer with a focus on digital access and communication, motivates several of the hypotheses, which follow below.

Table I: Number of API mashups, i.e. recombinations, and growth in market value by retailer. Mashups from Evans and Basole (2016), growth from Compustat (2017).

	Amazon	**Best Buy**	**Macy's**	**Sears**	**Target**	**Walgreen**	**Walmart**
Cumulative Mashups	329	9	1	3	3	7	1
10 Year Growth Rate	21,000%	411%	-31%	-91%	59%	284%	110%

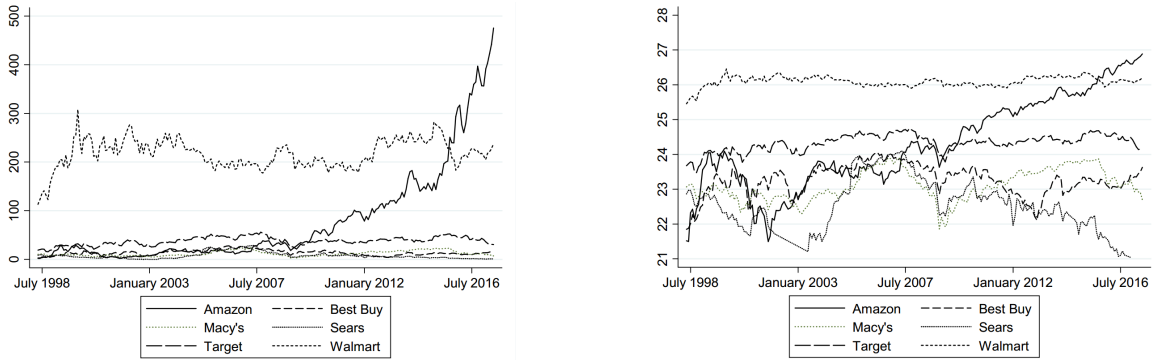


Figure 2: Market capitalization for Amazon, Best Buy, Macy's, Sears, Target and Walmart in (a) billions and (b) log scale, using Compustat (2017) data.

Hypothesis 1: API adoption increases the profitability and value of firms

Firms adopting an API strategy might see increases in profitability due to new products or new sales channels. For example, open APIs can facilitate development of complementary apps (Parker and Van Alstyne, 2017), while making it easier for customers to reach firms via mobile phones (Iyer and Henderson, 2010). Additionally, the tighter control an API gives firms over outside access could help firms price discriminate among existing digital products or sell new types of data services (Tiwana et al., 2010). APIs are more modular than traditional code, potentially increasing data and software access, reuse, and recombination (Yoo et al.,

2012; Baldwin and Clark, 2000). The potential to remix resources in new ways creates option value, which in turn encourages third parties to join the ecosystem (Baldwin and Clark, 2006). Therefore APIs should help firms get more effective labor from a certain expenditure on programmers. If programmers create investments that give benefits over time, then APIs should have benefits that manifest over time as well.

In our analysis, we proxy for a firm's API strategy adoption by whether the firm is observed using at least one API. We measure firms' long term profitability primarily through market value. These estimates face three main challenges.

First, the market value of a firm is a function of both current and expected productivity. Some of the anticipated value of API adoption is likely priced into the value of a firm's stock before they actually adopt the technology. This effect biases our estimates of the impact of API adoption on market cap towards zero.

Second, is the fact that we only observe API use for a subset of API users, not for firms outside our sample. Large, technologically savvy firms can create and maintain their own APIs using a variety of programming languages. Smaller firms, and ones for whom APIs are not a core competency, are more likely to seek the help of API management firms.[6] These firms help companies design and implement APIs to manage internal and external data flows and communications. Whether a firm designs their own APIs or partners with another API management firm is not reported in our data. If APIs increase market value, but our "control" group includes a mix of API adopting and non-adopting firms, this will also bias our estimates toward zero. This effect would be exacerbated if, as seems likely, firms with a core competency in APIs that design and manage them themselves see an even more positive effect from their use.

Third, and most challenging, is the problem of endogeneity. Firms that are already more successful, or anticipate growth, may be more likely to adopt APIs. More subtly, suppose that

[6]Some implement API proxies or cloud services. An API proxy is an intermediary, controlled by the API management company, through which calls to the customers APIs pass. These API proxies can provide additional functionality, including security (against hackers or DDOS attacks) or analytic tools. Some API management firms have designed computer languages and developer tools for creating APIs. One example is RAML (RESTful API Modeling Language), which Mulesoft has released under an open source license to encourage adoption.

some firms become well placed to adopt a platform strategy. Such firms might both adopt APIs and see increases in market value. We attempt to deal with endogeneity and other confounds through several different econometric specifications. That being said, our estimates are of the treatment effect on the treated group. They should not be interpreted as predicting the effect of API adoption on any random firm. Rather, we estimate the effect of API adoption on the types of firms who select into the treatment.

Hypothesis 2: Firms learn how to improve their API use over time

Any new technology takes time to implement. In reviewing the literature on how IT affects firms, Brynjolfsson and Hitt (2000) emphasize the importance of complementary investments - both organizational and technological – to make IT more productive. APIs by their nature are even more flexible and adaptable than earlier generations of IT. The variety of upstream (B2B), internal, and downstream (B2C) opportunities give rise to different API business cases for supply chain efficiencies, internal operations, and customer access channels. Musser (2013) argues that API implementation cannot be a cookie cutter process that simply formalizes existing data "cow-paths" within a firm. Rather, API designers must develop an appropriate company specific vision.

APIs can also facilitate the networking of disconnected pockets of expertise (Purvis et al., 2001), integration of new software into legacy software (Joseph et al., 2016) and speed IT deployment (Iyer and Subramanian, 2015). As a firm discovers how to take advantage of these possibilities it could become more innovative.

We judge whether firms are improving in their use of APIs in two main ways. First, we examine whether the impact of APIs on financial outcomes varies with time elapsed since date of adoption. Second, we explore whether firms make logical adjustments to their API use strategies in response to data breaches. These proxy management decisions over time.

Hypothesis 3: Open APIs allow firms to substitute away from internal R&D

APIs have two main orientations, open or closed, based on whether access is outward facing (open) or inward facing (closed). A closed API is only accessible to individuals or organizations working for the firm, which can strengthen security. The bulk of APIs are closed (Jacobson et al., 2011). Closed proprietary APIs are designed to enhance efficiency and internal agility.

For example, Amazon Web Services began as an internal project to reduce duplication of effort (Huckman et al., 2012). Hospitals, such as Cleveland Clinic, use internal APIs to share electronic medical records across different shifts and different teams of medical practitioners.[7] Verizon uses internal APIs to allow employees to access corporate services while making on-site service calls or activating new cable lines. To enable new tasks on old devices or old tasks on new devices, the firm simply creates a new app to talk to the API.

By contrast, open APIs may face upstream toward suppliers (B2B) or downstream toward consumers (B2C). Open APIs grant outsiders access to internal resources. They can also provide the infrastructure to support a partner ecosystem. By investing in an API developer portal by issuing "keys," a firm invites others to make complementary products. As a B2C example, Walgreens opened an API for photo printing services at its drugstores. Developers who incorporated Walgreens printing into their apps then drove users to print photos from their phones, social networks, and cloud accounts (Iyer and Subramanian, 2015). This externalized R&D increased sales both via photo services and in-store traffic.[8] In a B2B context, APIs can improve order entry and supply chain integration.[9]

Giving outside actors easier access through a developer portal means that outside partners will be more likely to develop complementary products. Opening an ecosystem through more permissive licensing, which is enabled by APIs, has been shown to increase complementary device development among handset manufacturers (Boudreau, 2010). Adding outside developers moves the boundaries of the firm (Parker et al., 2017), which been shown to improve performance (Chesbrough, 2006; Adner and Kapoor, 2010).

If closed APIs boost internal efficiency, their effect on R&D expenditures is potentially ambiguous as the firm enjoys both income effects and substitution effects on allocation of these savings. Open APIs, however, should have an unambiguous effect. They allow free outside

[7]Source: Bryan Kirchner, Google API Strategist, personal interview June 20, 2017.

[8]Other B2C uses of APIs include identification, billing, and service provisioning. On purchase of a mobile phone, for example, API services include porting a person's number, user authentication, SIM card assignment, phone activation, and account generation (Harvey, 2017).

[9]Ordering parts just-in-time or expanding cloud computing services on-demand has cut costs of ownership and improved work flow (Haranis, 2017). Insurance company Allstate packages analytics services based on driver risk scores, roadside assistance, and vehicle telematics for sale to rival insurers, auto manufacturers, and ride share services (Boulton, 2017). Uber and Lyft rely on Allstate data to vet applicant drivers.

developers, who previously played no role, to substitute for costly internal developers, leading a firm to spend less on its own R&D. On balance, we conjecture that open APIs substitute for R&D spending. In our analyses, we proxy for the openness of an API by observing the presence or absence of an external developer portal.

Hypothesis 4: APIs increase the risk of data hacks

The large purported benefits of API use raise the question of when *not* to use APIs. A principal answer is to avoid using open APIs for fault intolerant systems such as surgical robots and pacemakers. They should also not be used for high risk security applications such as military targeting and regional power grids.[10] In such cases, either experimentation is unwelcome or harmful intent is present. As a practical matter, a firm should not open APIs unless it is prepared to handle the increased traffic, unanticipated complications, and support requests that new third party interactions will precipitate.[11] Failed APIs result from failing to meet the service expectations of developer-consumers.

Several notable data breaches have been tied to flaws in APIs. In August 2018, T-Mobile announced an API data breach had exposed private data of more than 2.3 million users (Spring, 2018). A few months later, Google shut down Goolge+, its much maligned social networking venture, after revealing that the private data of more than 52 million users had been exposed to third parties through its APIs (Newman, 2018). Both of these breaches are examples of "leaky APIs". A leaky API is one that is vulnerable to hacking, misuse, or unintended disclosures because third parties are not properly metered or controlled when they request data.

The idea that an API may boost the efficiency of a firm, but at the risk of introducing a greater danger of data-breaches, is consistent with a growing literature on a performance-security trade-off, the Paradox of Exposure noted earlier. Ransbotham (2016) finds that open source software - while often functionally superior - is more likely than closed source software to have zero-day exploits. He argues that, although open code allows programmers to adapt it, accessibility also allows black-hats to identify avenues of attack. Kamiya et al. (2018) find that cyberattacks that leak personal financial information are associated with reductions in sales

[10]Source: Evans Data Corp.
[11]Source: Uri Sarid, CTO Mulesoft, interviewed June 21, 2017.

growth, investment, and stock market performance. Cumulative abnormal returns from one day before to one day after a breach are .84% while those from five days before to five days after are 1.1%. Successful cyberattacks also reduce CEO bonuses. The authors suggest boards perceive cyberattacked firms to have taken too many risks. This can create incentive problems if executives bear the risks of firms using APIs while shareholders reap the rewards.

Several other papers have also examined the effect of hack events on firm outcomes, some using data similar to ours. For example, Makridis and Dean (2018) also use data breach reports from the Privacy Rights Clearinghouse and match this data to Compustat financial data. They find a 10% rise in records breached is associated with a .2% fall in firm productivity. They note, however, that sample selection issues in publicly reported breach data makes precise identification of the treatment effect difficult. Spanos and Angelis (2016) perform a systematic literature review of the impact of information security events on stock market outcomes. They review 45 studies from 37 papers. Over 75% of these studies find a statistically significant effect of digital security events on stock prices. None of these studies, however, use API data to analyze technology adoption and usage behaviors per se.

We evaluate the hypothesis that API use can increase the risk of hacks in several ways. First, we estimate an event study model to see whether firms experience more hacks after adopting APIs. Certain specifications also include controls for measures of API security and openness. Then we investigate API flows in the months before and after the announcement of a data breach, explorations beyond the scope of prior work. Finally, we explore whether firms change their use of APIs after a data breach.

III. Data

Our analysis relies on merging three sets of data, one on firm level financial performance, one on API level data transfers, and one on firm level data breach events.

A. Data on API Use

Our main source of API usage data is a major API management firm. Companies that partner with this firm use its tools to create and operate APIs. The API management firm also helps partners regulate calls, data flows, and developer keys.

The API management firm provided us with monthly records of API use for 273 separate accounts. This includes the name of each API used, as well as the number of calls and bytes processed by each API in a given month. Data on calls processed by partner firms' APIs span December 2012 to September 2016. Data on bytes processed span December 2012 to May 2016. We designate the first date that we observe any call to any of a firm's APIs as the API adoption date. We succeeded in mapping these accounts to 123 publicly traded firms.

Figure 3 documents the total number of publicly traded firms we observe using APIs over time. It also displays the number of these firms we observe using a developer portal. A developer portal is a website giving documentation to the public about how a firm's APIs work. We collect this information using the Wayback Machine at the Internet Archive.[12] We use the presence of a developer portal as a proxy for the openness of a firm's APIs. Note that several firms had developer portals before we first observe them using APIs. This could indicate that they first tried to implement an API strategy without the help of an API tool provision company (or with the help of a competing one) or that they set up developer portals in advance of their APIs becoming operational. Firms adopted APIs smoothly over this interval.

Figure 4 reports the total number of APIs, API calls, and API bytes of data flow that we observe in each month. We have 2,453 firm-months of API usage data. The average firm has 160 million API calls in a given month, as well as 1.98 trillion bytes of data. The average firm has 31.4 APIs.

We categorize APIs by their purpose (e.g. login, search, payment, etc.) and orientation (e.g. B2B, internal, B2C). Details on how we do so can be found in the Appendix. Figures 5 and 6 report the log of the total number of calls processed by APIs as a function of the number of months since first API adoption. The panel is balanced, reporting calls for firms with at

[12] waybackmachine.org

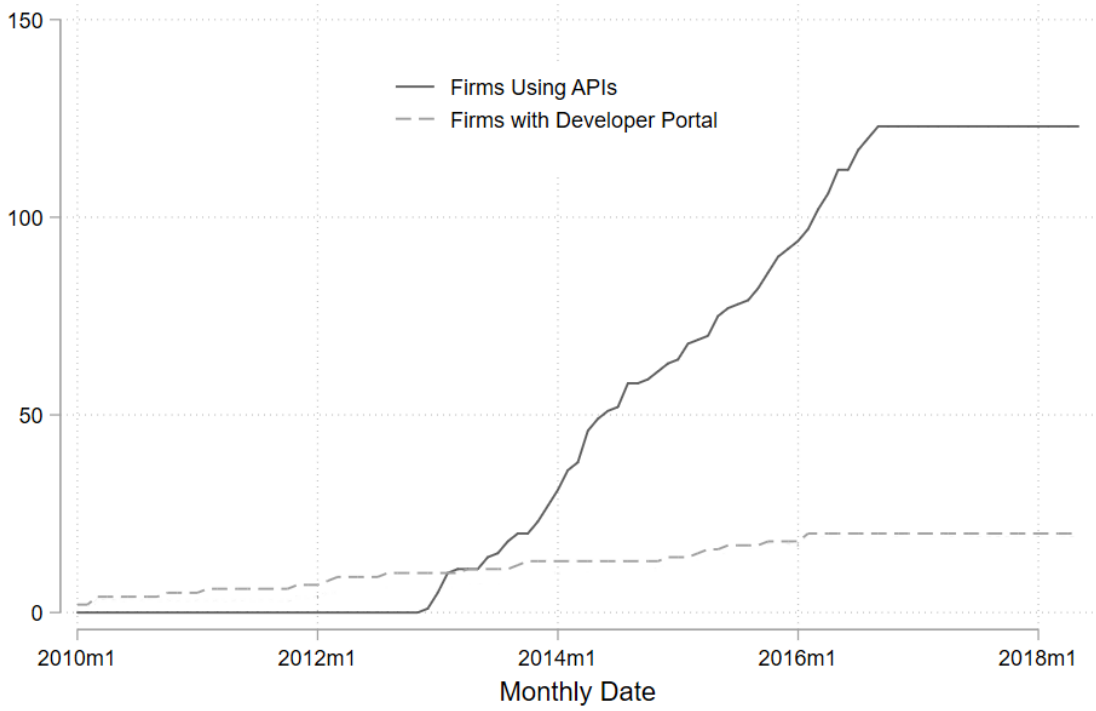


Figure 3: Number of public firms in our sample using APIs managed by our API tool provision company or who have a developer portal.

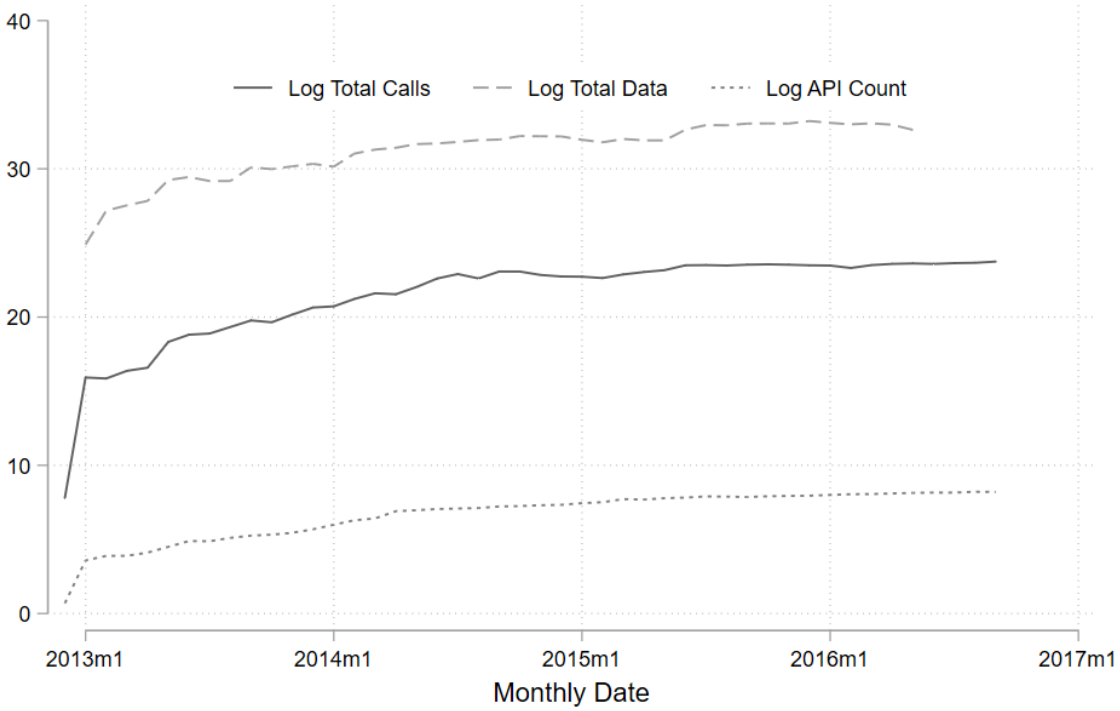


Figure 4: Total number of log calls, data, and APIs.

Table II: Total number of log calls, bytes, and APIs. Averages by firm-month. 2,453 firm-months of API call data.

	Mean	Std Dev	Max	N (Firm-Months)
Monthly Calls (Millions)	160	531	6,740	2,453
Monthly Data (Trillions of Bytes)	1.98	10.0	149	1,882
Number of APIs	31.4	46.2	433	2,453

least 12 months of observed API usage.

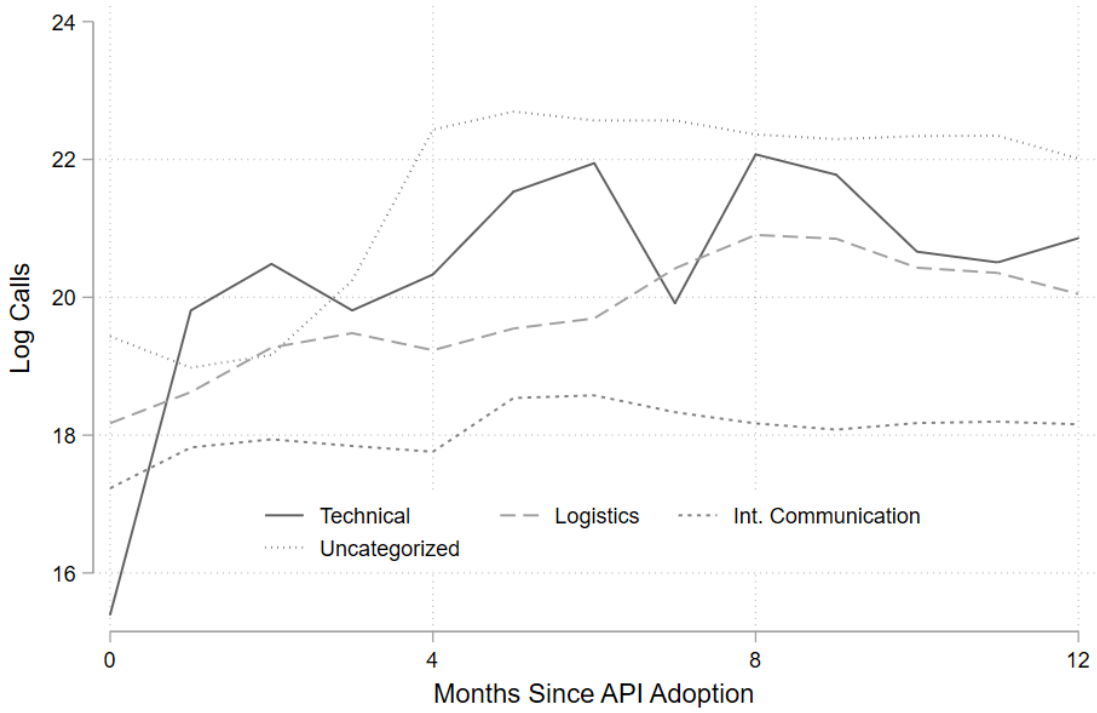


Figure 5: Calls of Most Commonly Used APIs by Purpose and Months Since First Adoption.

B. Complementary Data

Our primary financial data comes from Compustat, with resolution at the month and quarterly level. Of 273 initial accounts, 89 matched Compustat NA and 34 matched Compustat Global yielding a total of 123 firms. The remaining accounts corresponded to NGOs, government organizations, private companies, or represented different accounts at the same firm, e.g. a conglomerate, that needed to be merged.

Figure 6: Less Commonly Used APIs by Purpose and Months Since First Adoption.

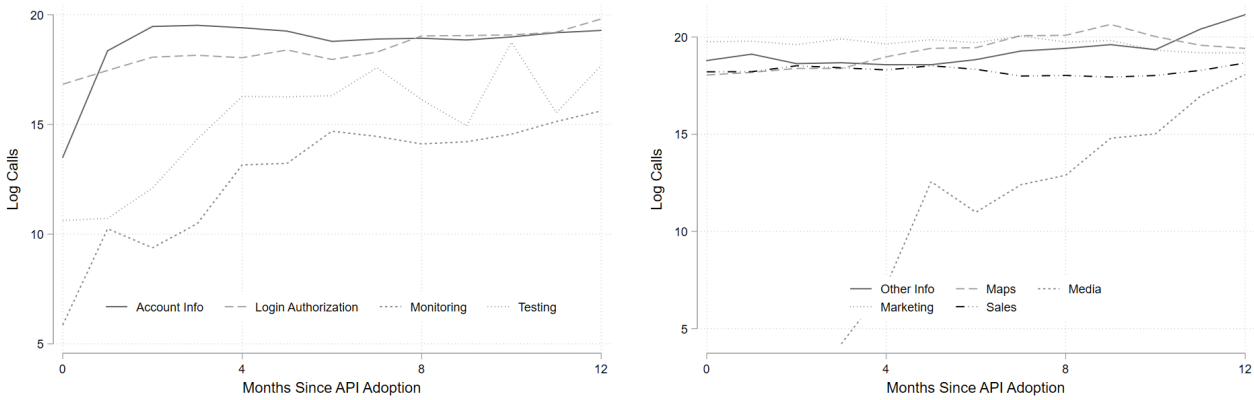


Table III reports basic financial information for these firms. They are quite large. Firms whom we observe using APIs had an average market value above $28 billion in a typical quarter.

To compare our API adopting firms with non-API adopters, we use a sample of 18,998 firms from Compustat NA whose API adoption is unknown. However, the key outcome variables (market value and RD expenditures) are not always reported for all of these firms over the entirety of the panel. This results in a panel sample of 12,849 firms for the market value regressions, and 6,829 for the R&D regressions, inclusive of the API adopting firms.

Table III: Financial characteristics of firms that adopt APIs during our sample period. All measures in millions of US dollars. All firm quarters pooled, 2008q1 through 2019q1.

	Mean	Std Dev	N (Firm-Quarters)
Market Value	28899.82	47854.1	3,159
R+D	134.3058	417.2701	5,831

Our final data source covers announced breach events recorded at the *Privacy Rights Clearinghouse* (PRC). This website reports the dates and characteristics of public breach announcements. We collect data for all public firms we observe using APIs back to 2008. PRC distinguishes six different breach event types: "PHYS", "PORT", and "STAT" events involve the theft of physical storage media, paper documents, and stationary devices. "INSD" events involve an insider. "DISC" events are unintended disclosures. "HACK" events are incidents

of hacking or malware leading to the data breach – these are of particular interest.[13] In our sample, 40 firms announced 189 data breach events since 2008, of which 65 were hacks.

The PRC also reports estimates of the number of records affected, which are typically included with the breach announcement. A majority of these, 121 of 189 data breach events, have such estimates. The distribution of records affected by data breaches has a long tail. The median number of records affected is 3578; the average is over 5 million. For data breaches without an explicit report, we estimate the number of records breached based on the announcement using the categories small, large, or massive. Based on moments of the known records breached distribution, we impute small data breaches as involving one record breached, large hacks as having 2037 records breached, and massive hacks as having 1.178 million records breached.

IV. APIs and Firm Financial Outcomes

Our first hypothesis is that API adopting firms will see faster growth than non-adopting firms. Figure 7 reports the distribution of market values for API adopters and non-adopters. API adopters are much larger than non-adopters. This may be due in part to APIs being an investment with a large fixed cost but low marginal cost to scale, making the investment more attractive for large firms.

Figure 8 reports the distribution of year on year growth rates for API using and non-using firms. Firms that use APIs see faster market value growth than non-adopters. This could indicate that APIs cause firms to grow faster than non-adopters but this is not dispositive. For example, larger firms may have generally grown faster during this interval, due to an increase in the prominence of "superstar firms" (David et al., 2017). However, to the extent that this is the case, the rise of fast growing superstar firms is itself partly due to the increasing importance of monopolistic platforms (Cusumano et al., 2019) and the APIs that sustain them. For both figures, we combine firms that will eventually adopt APIs with those that already have.

[13]PRC also records four "CARD" event types for firms in our data, which involve payment card fraud, and which are outside our interest.

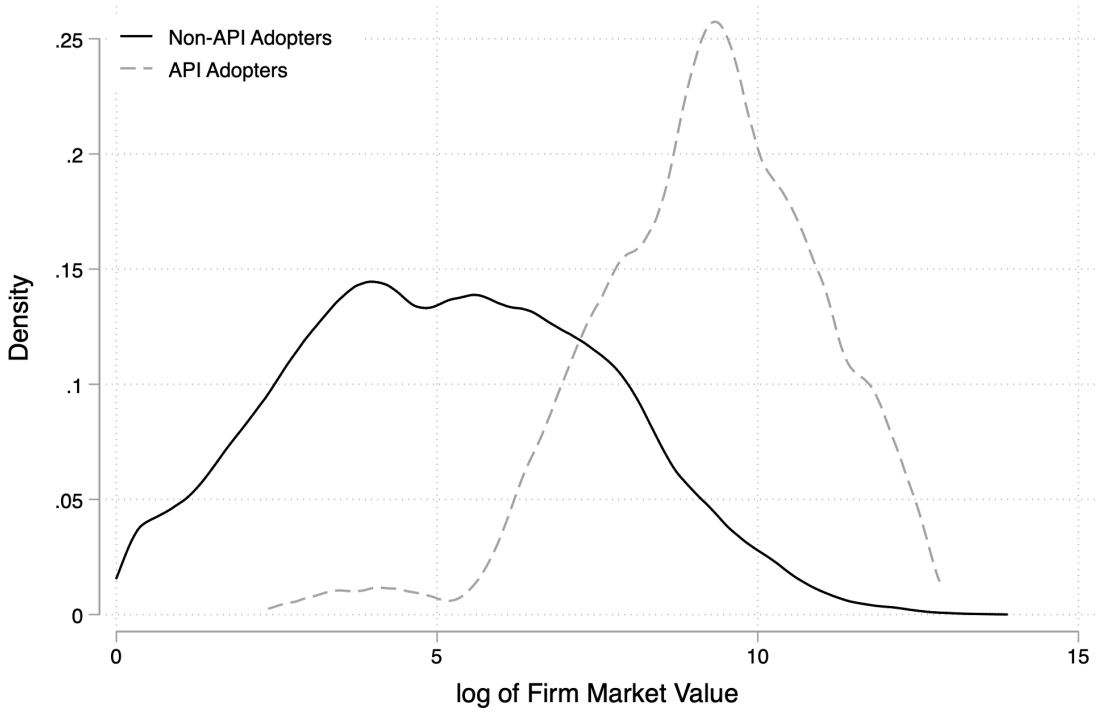


Figure 7: Density plot of log market value for publicly traded API adopters and non-adopters. Quarterly data 2008 through 2018.

A. Difference in Difference Estimates of Firm Financial Outcomes

The above figures are suggestive, but they do not imply a causal link between API adoption and market value growth. To estimate the role of APIs in firm growth, we estimate a series of leads-and-lags style difference-in-difference regressions where treatment is the timing of a firm's known adoption of APIs in our proprietary data. We define the date of API adoption as the first period a firm is observed using APIs.

Our estimating equations are

$$log(1 + MV_{i,t}) = \alpha_i + \tau_t + \sum \beta_j A_{i,t} + X_{i,t} + \epsilon_{i,t} \tag{1}$$

and

$$log(1 + R\&D_{i,t}) = \alpha_i + \tau_t + \sum_j \beta_j A_{i,j,t} + X_{i,t} + \epsilon_{i,t} \tag{2}$$

In equation (1) the outcome is the log of one plus market value, whereas in (2) the outcome is the log of one plus R&D expenditure. The equations are otherwise identical. In both equations,

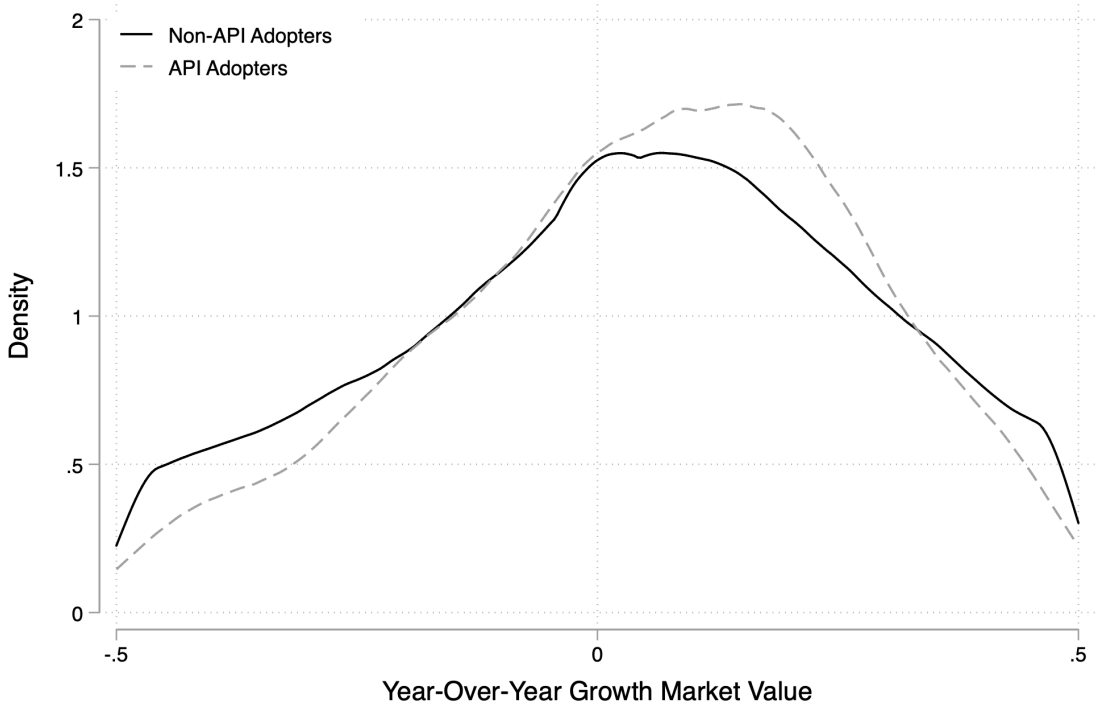


Figure 8: Density plot of year on year log market value change for publicly traded API adopters and non-adopters. Quarterly data, 2008 through 2018.

α_i is a firm fixed effect, τ_t is a quarter fixed effect and $A_{i,j,t}$ is a vector of dummies indicating the number of periods before or after a firm first adopted APIs. The dummies bin the quarters surrounding adoption into eight categories: yearly bins for the four years previous to adoption and two years post-adoption, and two additional bins for firms more than four years to adoption and more than two years post adoption.[14] Firms that we never observe using APIs take on the value $A_{i,j,t} = 0$ in all periods for all leads or lags j. $X_{i,t}$ are additional contemporaneous firm controls, which are included in certain specifications.

Estimates all control for a firm fixed effect. Therefore, they report the increase in a firm's market value, over a given horizon, after API adoption. Regressions all also control for a quarter fixed effect, meaning that the results are not driven by API adoption tending to occur in periods of rapid market value growth for all firms. A sufficient condition for β_j to be a causal estimate of API adoption on firm market value or R&D expenditures is that the timing of the adoption of APIs is uncorrelated with firm and time specific shocks to the outcome variable of interest.

[14]The period of adoption is included in the 0-1 years post adoption bin, so that bin runs from zero quarters post-adoption through three quarters post adoption. Subsequent bins include quarters four through seven, etc.

Table IV reports four variations on equation (1). The first column reports the results of a regression on the full sample, with quarter and year fixed effects, but no additional controls. This is our main specification. It detects a large and significant positive effect of API adoption on firm market value. They suggest that the average API adopter will see a 19.1% increase in market value after one year, and 20.1% increase in market value in subsequent years (H1). This growth in the size of the point estimates over different horizons is also consistent with firms learning how to better use their APIs over time (H2).

The second column adds additional controls for contemporaneous firm characteristics. While a firm fixed effect controls for any time-invariant factor that could influence both market value and API adoption, it remains possible that there is a time varying firm characteristic that connects the two. The controls are firm net goodwill, total revenues, debt, and operating expenses. The results of the regression are qualitatively similar. The fact that an effect on market value remains after controlling for these important firm attributes suggests that the effect of API adoption on market value operates through reducing non-operating expenses or through expectations about future increases in profitability. These would be different mechanisms underpinning H1.

The third column restricts attention to a balanced subset of firms. This specification drops any firms without observed market values at the beginning or end of the sample. Hypothetically, an unbalanced sample might bias estimates if all firms grow over time, but newly entering firms enter with a smaller size. However, the results are very similar in this variation.

The final column of Table IV restricts attention to firms that we also observe with a developer portal. Firms that adopt APIs, but are never observed with a developer portal, are dropped from the sample. This change restricts the number of observed API adopters dramatically. For this subset of firms the evidence is mixed. The point estimate of two or more periods post API adoption is even larger than in previous specifications and highly significant. The point estimate of the effect of four or more years prior to API adoption is even larger, but is not significant. In other words, while the effect of being more than two years post-API adoption is large and significant, it is not significantly different from being more than four years prior to API adoption.

Table IV: Estimates of the impact of API adoption on log market value following equation 1

	Main Model	Extra Controls	Balanced Panel	APIs & Dev Portal
	log Market Value	log Market Value	log Market Value	log Market Value
4 or more years until API adoption	-0.128	-0.00765	-0.143	0.399
	(0.129)	(0.139)	(0.126)	(0.334)
3-4 years until API adoption	-0.0551	0.0261	-0.0662	0.247
	(0.113)	(0.120)	(0.107)	(0.322)
2-3 years until API adoption	-0.0240	0.0458	-0.0513	0.293
	(0.101)	(0.110)	(0.0947)	(0.304)
1-2 years until API adoption	-0.0216	0.0490	-0.0454	0.218
	(0.0951)	(0.104)	(0.0872)	(0.307)
0-1 year until API adoption	0.0704	0.136	0.0598	0.179
	(0.0835)	(0.0902)	(0.0720)	(0.277)
0-1 year since API adoption	0.111	0.176*	0.0945	0.166
	(0.0719)	(0.0796)	(0.0577)	(0.243)
1-2 years since API adoption	0.191**	0.208**	0.172**	0.349+
	(0.0629)	(0.0651)	(0.0527)	(0.197)
2 or more years since API adoption	0.201***	0.247***	0.191**	0.321***
	(0.0584)	(0.0558)	(0.0632)	(0.0918)
Constant	5.319***	5.276***	5.282***	5.297***
	(0.0116)	(0.0135)	(0.0113)	(0.0116)
N	316738	278711	274356	314820
Firms	12849	11840	11672	12844
API Adopters	78	78	71	16
Firm FEs	Yes	Yes	Yes	Yes
Quarter FEs	Yes	Yes	Yes	Yes
Balanced	No	No	Yes	No
Additional Controls	No	Yes	No	No
Only Firms w/ Dev Portals	No	No	No	Yes

Notes: Standard errors in parentheses, clustered at the firm level. Additional controls include control for firm level net goodwill, total revenue, long term debt, and operating expenses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table V estimates equation 2 , which uses firm $R\&D$ expenditures as the outcome of interest. The layout of columns matches that of Table IV. The first column only includes quarter and year fixed effects, the second column includes a vector of additional controls - firm net goodwill, total revenues, debt, and operating expenses - the third column restricts focus to a balanced panel of firms, and the final column drops firms that adopt APIs without also having a developer portal.

The different restrictions give qualitatively similar results. API adopters, in the years before API adoption, spend significantly more than the average firm on R&D. However, after two years post adoption the estimates are tightly estimated nulls, indicating that API adoption leads high R&D expenditure firms to reduce their expenditure to average levels (H3).

Consistent with theories that APIs "invert the firm," moving value creation from inside to outside (Parker et al., 2017), API adoption leads to a decrease in $R\&D$ expenditure despite gains in market value. Focusing on the main specification – Column 1 – API adopting firms spend 25.3% more than non-adopters four or more years before API adoption. Two years after API adoption, they spend .2% less, essentially indistinguishable from average firms. Since the average firm in our sample spends 134.3 million on R&D expenses, a 25.3% decrease represents spending 33.98 million less on R&D. While there are few observations, the decrease is the most dramatic for firms that use both APIs and developer portals. These firms reduce their spending from 31.5% above average in the third year before adoption to average levels two years after adoption. Results support H3 that firms substitute external for internal R&D.

Across Tables IV and V, the size and significance of this section's estimates are particularly surprising given that estimates of the effect of API adoption face three strong headwinds that bias estimates of coefficients toward zero.

The first headwind is that our "untreated" observations are contaminated, in the sense that some of the firms in our control group use APIs. Mixing treated and untreated observations would tend to bias coefficient estimates toward zero.

A second headwind, market anticipation, suggests that the effect of API adoption on market value should not be precisely timed to the date of first API adoption. If stockholders anticipate a profitable API implementation then the market value of a stock should increase before

Table V: Estimates of the impact of API adoption on log R&D following Equation 2

	Main Model	Extra Controls	Balanced Panel	APIs & Dev Portal
	log R&D	log R&D	log R&D	log R&D
4 or more years until API adoption	0.253***	0.325***	0.195**	0.498**
	(0.0644)	(0.0695)	(0.0684)	(0.161)
3-4 years until API adoption	0.246***	0.287***	0.178**	0.315*
	(0.0535)	(0.0609)	(0.0571)	(0.146)
2-3 years until API adoption	0.194***	0.223***	0.138*	0.247+
	(0.0505)	(0.0536)	(0.0547)	(0.146)
1-2 years until API adoption	0.184***	0.217***	0.128*	0.261*
	(0.0521)	(0.0525)	(0.0580)	(0.132)
0-1 year until API adoption	0.154***	0.149**	0.108*	0.239*
	(0.0436)	(0.0500)	(0.0490)	(0.104)
0-1 year since API adoption	0.0928*	0.114*	0.0672	0.139+
	(0.0362)	(0.0448)	(0.0418)	(0.0783)
1-2 years since API adoption	0.0586	0.0996*	0.0670	0.0962
	(0.0367)	(0.0434)	(0.0412)	(0.101)
2 or more years since API adoption	-0.0205	0.0103	0.00288	0.0182
	(0.0348)	(0.0329)	(0.0374)	(0.0888)
Constant	1.319***	1.298***	1.206***	1.309***
	(0.00945)	(0.0107)	(0.00925)	(0.00944)
N	141484	130113	111521	140390
Firms	6829	6803	5647	6824
API Adopters	57	57	42	14
Firm FEs	Yes	Yes	Yes	Yes
Quarter FEs	Yes	Yes	Yes	Yes
Balanced	No	No	Yes	No
Additional Controls	No	Yes	No	No
Only firms w/ Dev Portals	No	No	No	Yes

Notes: Standard errors in parentheses. SEs clustered at the firm level. Additional controls include control for firm level net goodwill, total revenue, long term debt, and operating expenses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

implementation. Similar anticipation could also drive an ex-ante implementation effect of API adoption on R&D expenditure. If a firm anticipates outside complementors will soon create free innovations, they may begin winding down R&D expenditure well in advance of API implementation.

The final headwind is that the full effects of API implementation may only make themselves felt over time. Large complementary investments and long lags in measurement are often implicated in delays associated with realizing the benefits of new technologies. Brynjolfsson et al. (2018) discuss how these lags can lead to macroeconomically important mismeasurement of the productivity effects of technological change. As our regressions reveal, the impact of API adoption increases over time.

Figures 9 and 10 plot coefficients estimated from regressions of financial outcomes on API adoption having occurred in a given number of quarters adjacent to adoption. The specification is identical to that of the first column of tables IV and V except the bins for time since adoption are quarters instead of years.[15]

As in the previous tables, outcomes are log market value and log R&D expenditure. Each figure displays the coefficients from two regressions – one restricted to firms with developer portals, and one without. Each of the four regressions includes 25 total leads and lags of API adoption. 95% confidence intervals with firm-clustered standard errors are also displayed.

Consistent with our previous regressions, there is a strong increasing trend in market value and decreasing trend in R&D expenditures for firms around the date of API adoption. The decrease in R&D is particularly pronounced for firms that also have a developer portal. However, consistent with our 'headwinds', the effect is not precisely specific to the quarter after adoption. An ex-ante effect from API adoption seems present. The evidence of this section overall is consistent with APIs having strong financial outcomes for firms, but an effect which firms and markets partly account for in advance, and get better at harnessing in the years after adoption.[16]

[15]Table X in the appendix reports estimates of column one specifications using quarter bins instead of years.

[16]Ideally in a setting like this an instrument for API adoption would be available. However, no clear candidate was identified. Discussants have proposed using inverse propensity score weighting. This approach overweights adoptions that are statistically unlikely to deal with endogeneity (Hirano et al., 2003). Others have proposed using previous adoption of APIs within an industry as an instrument for future adoption. If previous adoption in an industry predicts future adoption, this would effectively overweight *more* likely adoptions in the analysis. With no clear instrument, our analysis proceeds as best possible.

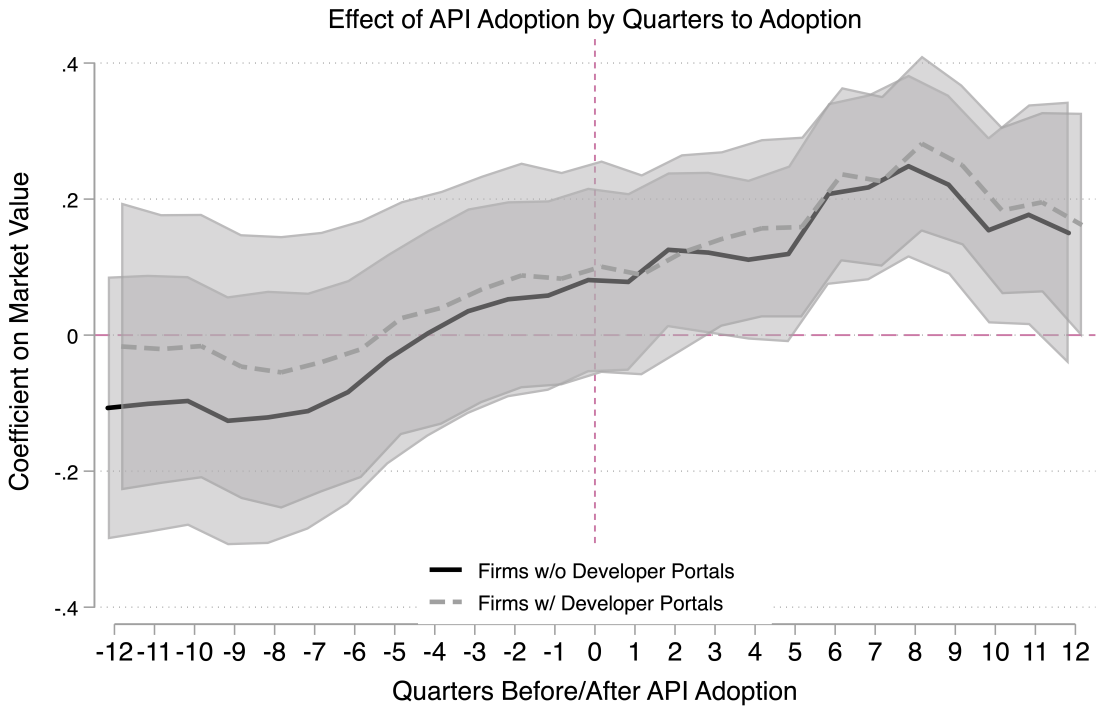


Figure 9: Effect of API Adoption on log Market Value for Firms with and without Open Developer Portals. Coefficient plot of estimates from two regressions where log market value is a function of different numbers of quarters pre and post API adoption. Unbalanced panel with 95% confidence intervals and firm-level clustered standard errors.

V. API Usage and Data Breaches

If, as the previous section suggests, evidence is consistent with APIs having a significant causal impact on business profitability, then why do not all firms adopt them?

Here, we investigate the prospect that API use, while boosting average firm efficiency, also increases risks of a data breach. We find that firms opening APIs do see an increased risk of hack events and unsophisticated firms lose more records when regressions control for the sophistication of API strategy employed. Breach risk worsens if the firm has an open developer portal.

We then turn our attention to how firms modify their use of APIs in response to hack and data breach events. We show that in the month before a hack is announced, firms significantly reduce data flows associated with uncategorized APIs. This likely corresponds to firms clamping down on data flows that are related to the breach. In the year after a data breaches, firms

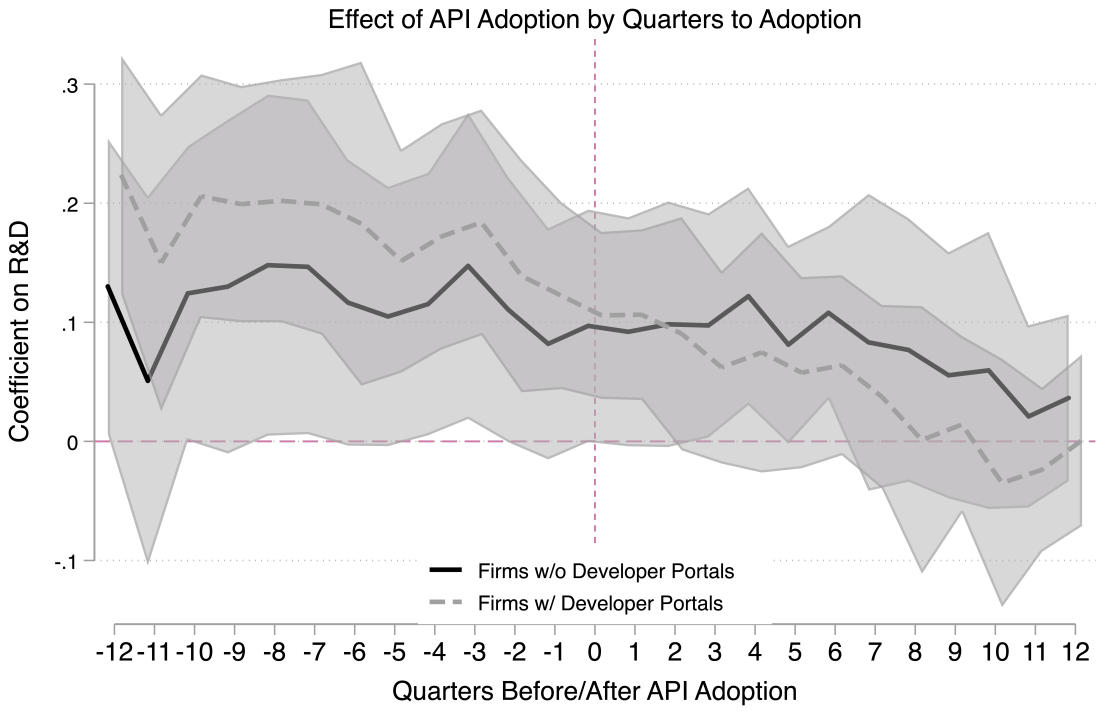


Figure 10: Effect of API Adoption on R&D Spending for Firms with and without Open Developer Portals. Coefficient plot of estimates show two regressions where log R&D expenditure is a function of different numbers of quarters pre and post API adoption. Unbalanced panel with 95% confidence intervals and firm-level clustered standard errors.

respond in appropriate ways depending on the type of breach.

The types of adjustments are intuitive. Firms hacked from online sources make increased use of security APIs (login and account authorization) in the subsequent year. Firms that lost physical storage media substitute toward digital media and increase their use of APIs for storing and transmitting data. The fact that APIs respond so dramatically and intuitively is also further evidence of the idea that these are important causes of data breaches. These are consistent with both H3 and H4.

Overall the evidence of this section is consistent with the idea that APIs, when implemented inexpertly, can increase hacking risk. But it also supports the idea that firms can adjust their use of APIs to minimize the types of breach risks they face.

A. Impact of API Opening on External Actor Hack Events

Our first question is whether a firm's decision to adopt APIs increases the likelihood or severity of cyberattacks. We measure severity based on the number of records reported as breached. For this analysis, we restrict attention to data breaches caused by outsiders. Cyberattacks correspond to "hacks" in the PRC taxonomy and represent malicious data intrusions and malware that result in data breaches.

For this estimate, we organize our firm and hacking events data at the monthly level. We estimate the likelihood of a cyberattack (or the number of records breached) as a function of how long it has been since a firm has adopted APIs, as well as measures of a firm's API sophistication and firm and year fixed effects. Depending on specification, we estimate a linear probability model as follows:

$$y_{i,t} = \alpha_i + \tau_y + \beta \mathbb{1}(0 \leq A_{i,t} \leq 3) + \eta \mathbb{1}(0 \leq A_{i,t} * DevPortal \leq 3) + \epsilon_{i,t} \tag{3}$$

where y_{it} measures either the occurrence of a hack event in a given month, or the log of one plus the count of records breached in hack events. α_i is a time-invariant firm fixed effect to control for firm-specific likelihood of hacking events.[17] Estimates also include a year fixed effect τ_y. We restrict ourselves to year fixed effects rather than quarter fixed effects, because we have only 65 hack observations. The parameter of main interest is β. It is a dummy variable equal to 1 if a firm opened up to APIs within the last 3 years. η is also of interest. It is a dummy variable equal to 1 if a firm has both opened up an API in the last three years and has a developer portal.

In some specifications, we also include a vector of controls for the sophistication and security of a firm's APIs $S_{i,t}$. We consider three proxies for a firm's API security. The first is a raw count of the number of APIs a firm has. A firm that has many smaller and more modular APIs is thought to be more resistant to cyberattacks than one having larger more general purpose APIs.[18] The second is the share of APIs that are devoted to testing. The third measure of API

[17] This also controls for industry effects, if firms in certain industries, say finance, are more likely to be targeted.

[18] Op. cit. Berlind

sophistication is the share of API data flows going to testing APIs.

Tables VI and VII report the results of estimating equation 3 with different controls. The point estimate of the effect of API adoption on hacks is .0119 (Table VI Column 1) although this estimate is only significant at the 10% level. Switching to the log of the number of records hacked as the outcome of interest, the estimated effect of API adoption is large and significant. We find that API adoption leads to a 17.6% increase in the number of records breached.

The effect of API adoption on hacks is stronger and more significant after controlling for measures of API security and sophistication. As anticipated, Table VII shows the number of APIs used, the share of APIs used for testing, and the share of data flows in testing APIs strongly negatively predict the risk and severity of hacks across almost all specifications. For example, Column 4 in both tables displays a specification where hack probability (Table VI) and severity (Table VII) is a function of API adoption and the share of API data used for testing APIs. The point estimates suggest that a firm increasing the share of API flows devoted to testing by 50% would reduce the risk of hacks by 1.8% and the number of records breached by 40.5%. The point estimate of the effect of opening a developer portal on the risk and severity of hacks is positive but only marginally significant.

Table VI: Impact of API Adoption on Hack Events

	Hack Event	Hack Event	Hack Event	Hack Event	Hack Event
Years 0 - 3 post API adoption	0.0119^{+}	0.0155*	0.0119^{+}	0.0121^{+}	0.0103
	(0.00671)	(0.00726)	(0.00664)	(0.00674)	(0.00644)
API Count (100s)		-0.00815^{+}			-0.00811^{+}
		(0.00478)			(0.00468)
API Count Testing Share			0.0000541		0.000294
			(0.000506)		(0.000695)
API Data Flow Testing Share				-0.0360**	-0.0487
				(0.0111)	(0.0388)
Developer Portal Years 0-3					0.0267^{+}
					(0.0154)
Year FEs	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes
R2	0.00607	0.00754	0.00607	0.00616	0.00939
Obs	5788	5788	5788	5788	5788
Firms	123	123	123	123	123

Standard errors in parentheses, clustered at the firm level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table VII: Impact of API Adoption on log(1+Records Breached)

	Data Hacked	Data Hacked	Data Hacked	Data Hacked	Data Hacked
Years 0 - 3 post API adoption	0.176*	0.229*	0.192*	0.180*	0.228*
	(0.0860)	(0.0952)	(0.0866)	(0.0859)	(0.0934)
API Count (100s)		-0.123*			-0.125*
		(0.0597)			(0.0593)
API Count Testing Share			-0.0119**		-0.0108^{+}
			(0.00434)		(0.00571)
API Data Flow Testing Share				-0.810***	-0.312
				(0.224)	(0.304)
Developer Portal Years 0-3					0.0950
					(0.185)
Year FEs	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes
R2	0.00858	0.0107	0.00926	0.00888	0.0117
Obs	5788	5788	5788	5788	5788
Firms	123	123	123	123	123

Standard errors in parentheses, clustered at the firm level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B. The Immediate Impact of Hacks On API Flow

If APIs enable data breaches, then data flows across the APIs should change around the time of the breach. In the case of large data breaches, this might show up as a spike in traffic. Alternatively, a hack might lead firms to restrict data flows to stem losses. This would lead us to detect a drop in traffic.

To understand how API flows vary in the periods before and after hacks, we estimate equations of the form

$$y_{i,t} = \alpha_i + \tau_t + \sum_{j=-6}^{6} (\beta_j Hack_{i,j,t}) + \epsilon_{i,t} \tag{4}$$

where y_{it} measures log of one plus API calls or data flows of a specified API type. $Hack_{i,j,t}$ is a vector of dummies indicating the number of months before or after a hack announcement. Coefficients β_j indicate how API flow varies before and after a hack announcement. α_i and τ_t are firm and month fixed effects.

Figure 11 reports the change in log API calls as a function of the number of months before or after a data breach is announced, including fixed effects. It plots estimates of equation (4), with points corresponding to estimates of the effect of being j months before or after the announcement. Figure 11 shows a "data cliff," clearly indicating a large decrease in calls and data flows in the month before the hack is announced. Note also that the *variance* of API call and data flows increases dramatically in the period before and during a hack announcement. This likely corresponds to some firms successfully closing hacked APIs, while other firms, unaware they have been hacked, inadvertently allow an increase in unauthorized disclosures.

In Figure 12, we restrict attention to APIs that are most reduced in the month before a hack is announced – those non-standard APIs whose purpose we could not classify. The fact that the total reduction in calls is driven by uncategorized APIs is suggestive of the fact that these are both the most vulnerable and poorly implemented of API types.

Figure 13 displays the same analysis for two other API types, Internal Communications and Testing respectively. As can be seen, there is a clear increase in use of testing APIs and decrease in use of internal communications APIs around the timing of the attack. These

Figure 11: Total Log API Calls (left) and Data (right) by months before and after hack. Coefficients estimated using equation 4 and show 95% confidence intervals.

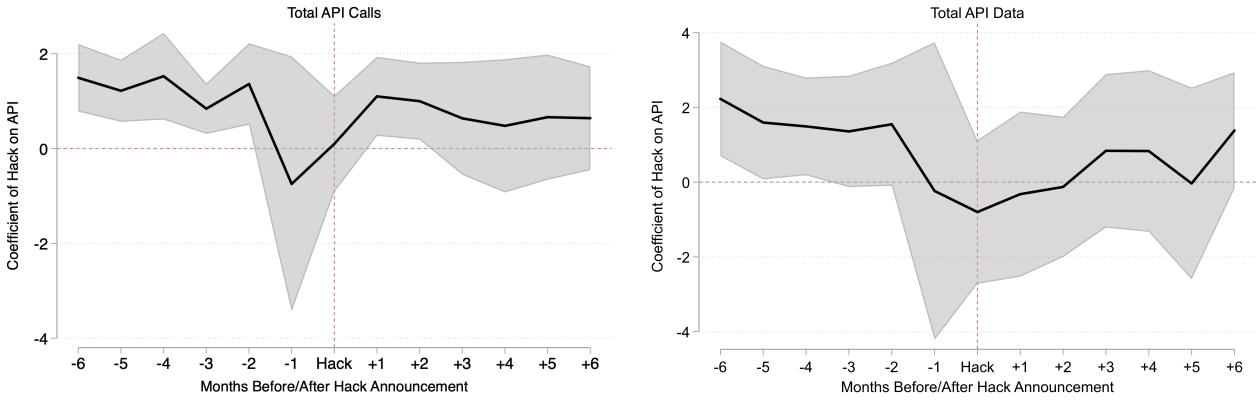
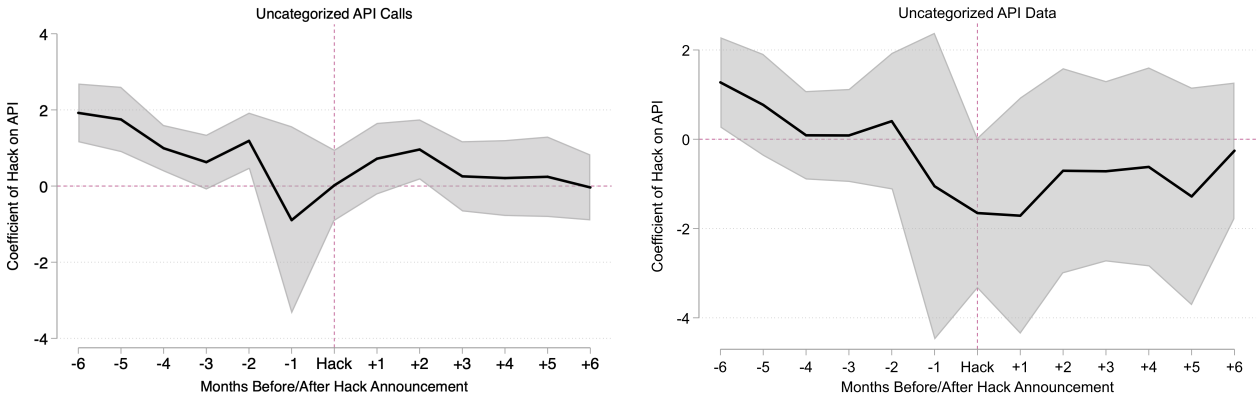


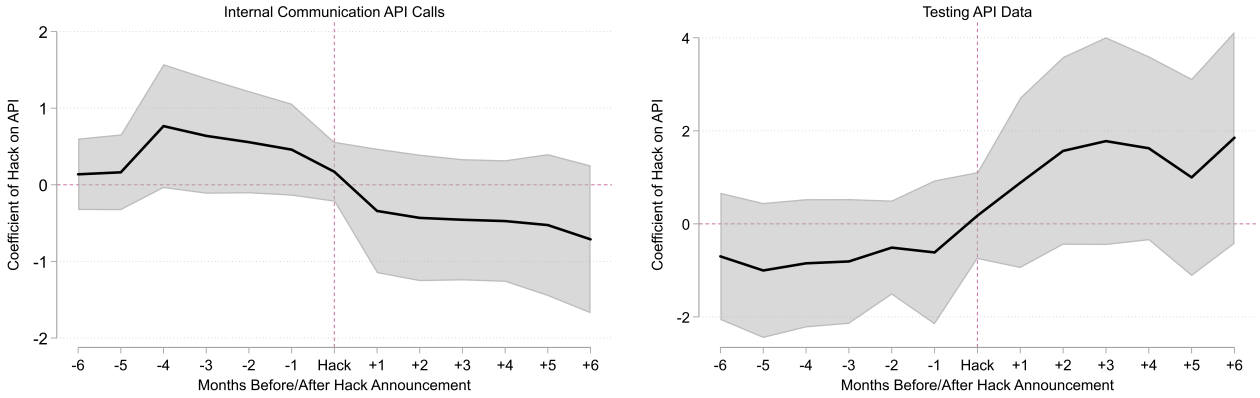
Figure 12: Uncategorized API Calls (left) and Data (right) by months before and after hack. Coefficients estimated using equation 4 and show 95% confidence intervals.



are logical responses to a data breach. If firms were using internal communications APIs for sensitive information, it makes sense for them to reduce this usage in the wake of a compromised IT system and rely on other means for communicating confidential messages. In response to a hack, firms also resort to API testing. This supports the idea that the use of testing APIs can help reduce the likelihood and severity of hacks.

Figures 17 and 18 in the appendix repeat this estimate for APIs of all types.

Figure 13: Internal Communication API calls (left) and Testing API data flows (right) by months before and after hack. Coefficients on months near hack as estimated by equation 4 with 95% confidence intervals.



C. Long Term Impact of Data Breaches on API Use

Given our evidence that APIs are related to hack events in a causal way, it is interesting to investigate how firms adapt their long term use of APIs in response to a data breach. Because of the richness of our data, we can follow what happens internally to API calls and data transfers. Specifically, we can investigate which APIs firms use more or less intensively following a security breach. We begin with a focus on hacks in particular.

We estimate a series of fourteen equations of the form:

$$log(1 + APIDataFlow_{fit}) = \alpha_i + \tau_t + \beta_f HackEventPastPeriod_{i,t} + \epsilon_{i,t} \quad (5)$$

where the dependent variables, $log(1 + APIDataFlow_{fit})$ are the volume of API flows of function f measured as the log of one plus the number of calls or log of one plus the bytes of data flow. τ_t is a monthly fixed effect and α_i is a firm fixed effect.

Tables VIII and IX show the result of estimating these fourteen equations. These tables report the change in API flows in the one and two year periods following a hack event respectively. Figures 14 and 15 present these same results in graphic form. These figures summarize the tables' regression results by plotting the magnitude and confidence intervals for β_f coefficients for each type of API function f. The first reports effects for the first year

following the hack event, and the second reports the effect in the second year.[19]

Total API calls increase in the year after a hack with a coefficient of 0.643, but this effect is not significant. The increase in API calls after a hack is driven by increases in the use of security (e.g. testing and logins) and operational (e.g. sales) APIs. The effect is particularly strong for testing APIs. We find test calls and data more than double following hack events, both significant at the 0.001 level. This suggests that firms suffering a breach might not have sufficient testing protocols and respond to external hacks by increasing their use of APIs designed for this purpose.

Column 2 of Table VIII shows that the volume of data through uncategorized APIs *decreases* in the year after a breach, although again not significantly. This result, alongside the fact that firms reduce their use of uncategorized APIs immediately before hack events are announced, suggests that uncategorized APIs are involved in the hacks themselves. The fact that these APIs used non-standard naming conventions would be unsurprising if these APIs were shoddily built (or intentionally built with back doors). In addition, Table IX shows a negative point estimate on the use of APIs classified as Technical or Internal Communication in the year post hack. After a data breach, it is unsurprising that firms trust their internal communications less out of concern for data interception.

Over the longer term, Figure 15 reveals that firms increase their overall use of APIs in the two years following a hack. This increase is driven by functional APIs, such as logistics, marketing, and other information, yet even involves uncategorized APIs. In effect, they appear to embrace the benefits of APIs after securing them from further attack.

Up to now we have focused exclusively on what happens to a firm after a hack. However, there are other types of security breaches whose events may impact API use. The Privacy Rights Clearinghouse defines five other types of security breaches of interest. We estimate the

[19]Table XI, in the Appendix, reports the same estimations with log of data flows as the outcome.

Table VIII: Impact of Hack Event on log API Calls by API Function

	Total API Calls (logs)	Un-categorized	Account Info	Login Autho-rization	Monitoring	Testing	Other Info
One Year Post Hack Event	0.643	-0.360	0.584	0.597	-0.0511	2.076**	0.193
	(1.063)	(0.847)	(1.907)	(0.900)	(0.172)	(0.793)	(1.044)
R2	0.259	0.114	0.147	0.0956	0.0903	0.169	0.150
One to Two Years Post Hack Event	1.122***	0.737**	0.355	0.570	-0.120	0.224	0.783+
	(0.227)	(0.246)	(0.478)	(0.369)	(0.175)	(0.334)	(0.422)
R2	0.287	0.130	0.147	0.102	0.0909	0.154	0.155

Notes: Standard errors in parentheses, clustered at firm level. Regressions include month and firm fixed effects, and all reflect 123 firms and 2535 observations (firm-months) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table IX: Impact of Hack Event on log API Calls by API Function (continued)

	Internal Comm	Logistics	Maps	Media	Marketing	Sales	Technical
One Year Post Hack Event	-1.227	1.413+	0.257	0.0259	0.807	0.915	-0.150
	(0.837)	(0.739)	(0.525)	(0.304)	(0.533)	(0.824)	(1.092)
R2	0.112	0.0731	0.0963	0.103	0.163	0.209	0.153
One to Two Years Post Hack Event	0.0650	0.635**	0.0753	0.189	0.671+	-0.178	0.271
	(0.285)	(0.229)	(0.275)	(0.261)	(0.364)	(0.350)	(0.324)
R2	0.100	0.0663	0.0963	0.104	0.164	0.205	0.155

Notes: Standard errors in parentheses, clustered at firm level. Regressions include month and firm fixed effects and all reflect 123 firms and 2535 observations (firm-months) $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

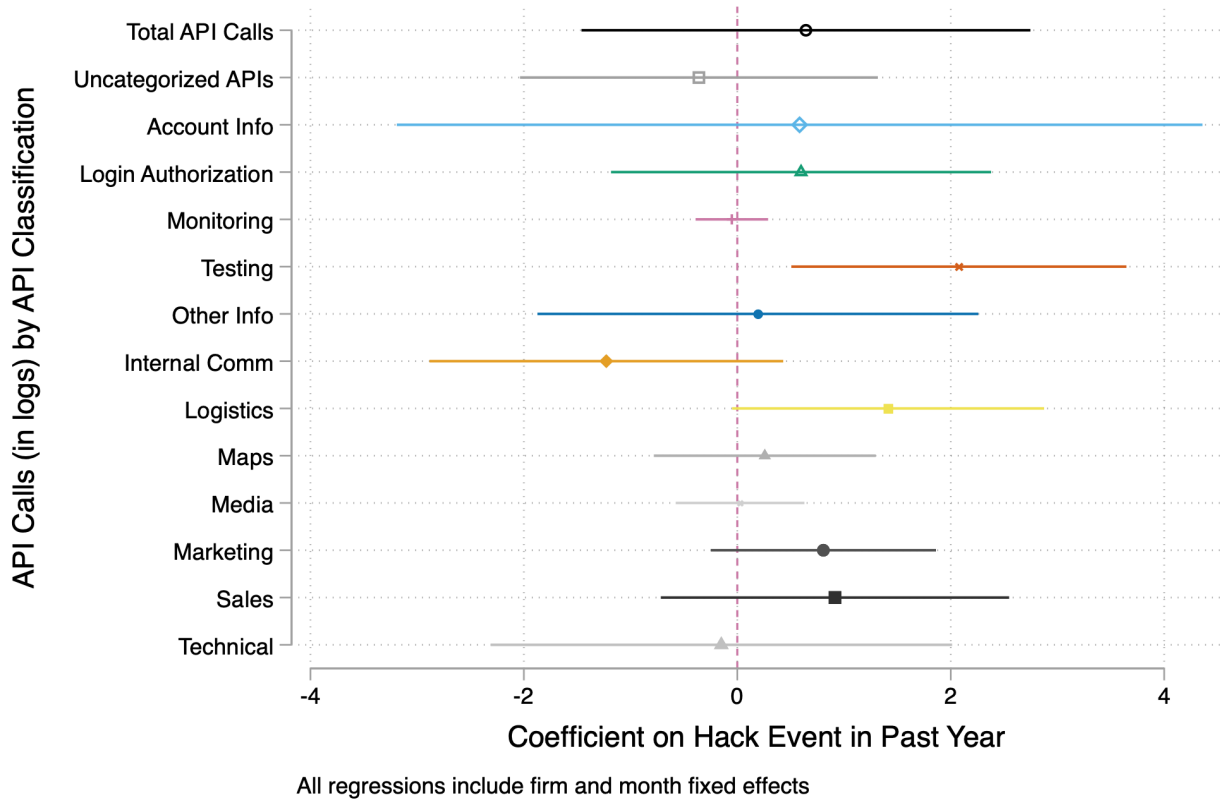


Figure 14: Effect of hack events in the past year on API Usage Intensity (log API Calls) by API Type. Each plotted point is the output from a separate regression where the outcome variables are API calls of a given type, as shown in tables VIII and IX. All regressions include month and firm fixed effects, clustered at the firm level, and show 95% confidence intervals. Testing doubles while internal communications fall.

following equation to test how other types of security breaches impact API use.

$$
\begin{aligned}
log(1 + APIDataFlow_{fit}) = \alpha_i + \tau_t \\
&+\delta_f UnintendedDisclosure_{it} + \zeta_f HackingEvent_{it} \\
&+\eta_f InsiderBreach_{it} + \theta_f PaperDocumentLost_{it} + \\
&+\lambda_f PhysicalStorageLost_{it} + \mu_f StationaryDeviceLost_{it} \\
&+\epsilon_{i,t}
\end{aligned}
\tag{6}
$$

In equation 6 the coefficients of interest are δ_f, ζ_f, η_f, θ_f, λ_f, and μ_f. We define the breach events similarly to the way we defined them for hack events, as binary dummy indicators if the

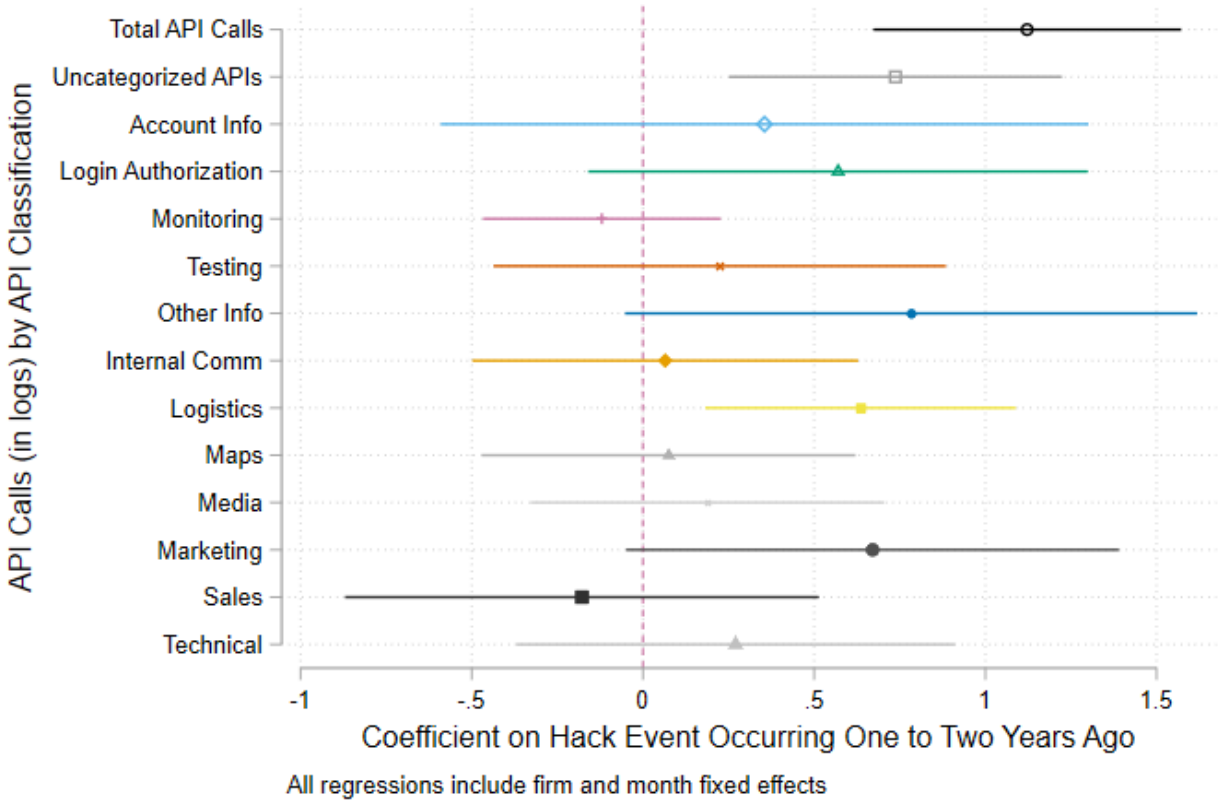


Figure 15: Regression Summary Coefficient Plot: Impact of Hack events occurring one to two years ago on API Use Intensity (log API Calls) by Classification of API. Each plotted point is the output from a separate regression where the outcome variables are API calls of a given type, as shown in tables VIII and IX. All regression include month and firm fixed effects, and show 95% confidence intervals. Total API calls rise.

firm experienced a breach event of a given type in the previous twelve months. This presents a multiplicity of equations to estimate, and for each we are interested in the magnitude and significance of these breach event dummies, which will tell us how APIs of a given type respond to breach events of a given type. Data flow, the outcome of interest, is measured as the log of one plus the amount of API calls or amount of data flowing through the API.

Estimates of Equation 6, with usage displayed for both calls and data, for a one year period after the data breach, are shown in Tables XI and XII in the Appendix. To aid interpretation, we also plot the coefficients and their standard errors for the API types of greatest interest in Figure 16. In each panel, the vertical axis shows the type of breach, and the horizontal axis shows the increase (or decrease) in use of that type of API in the the year following the breach.

Several points bear mentioning. 1) Hacking by outside parties leads to a doubling of API

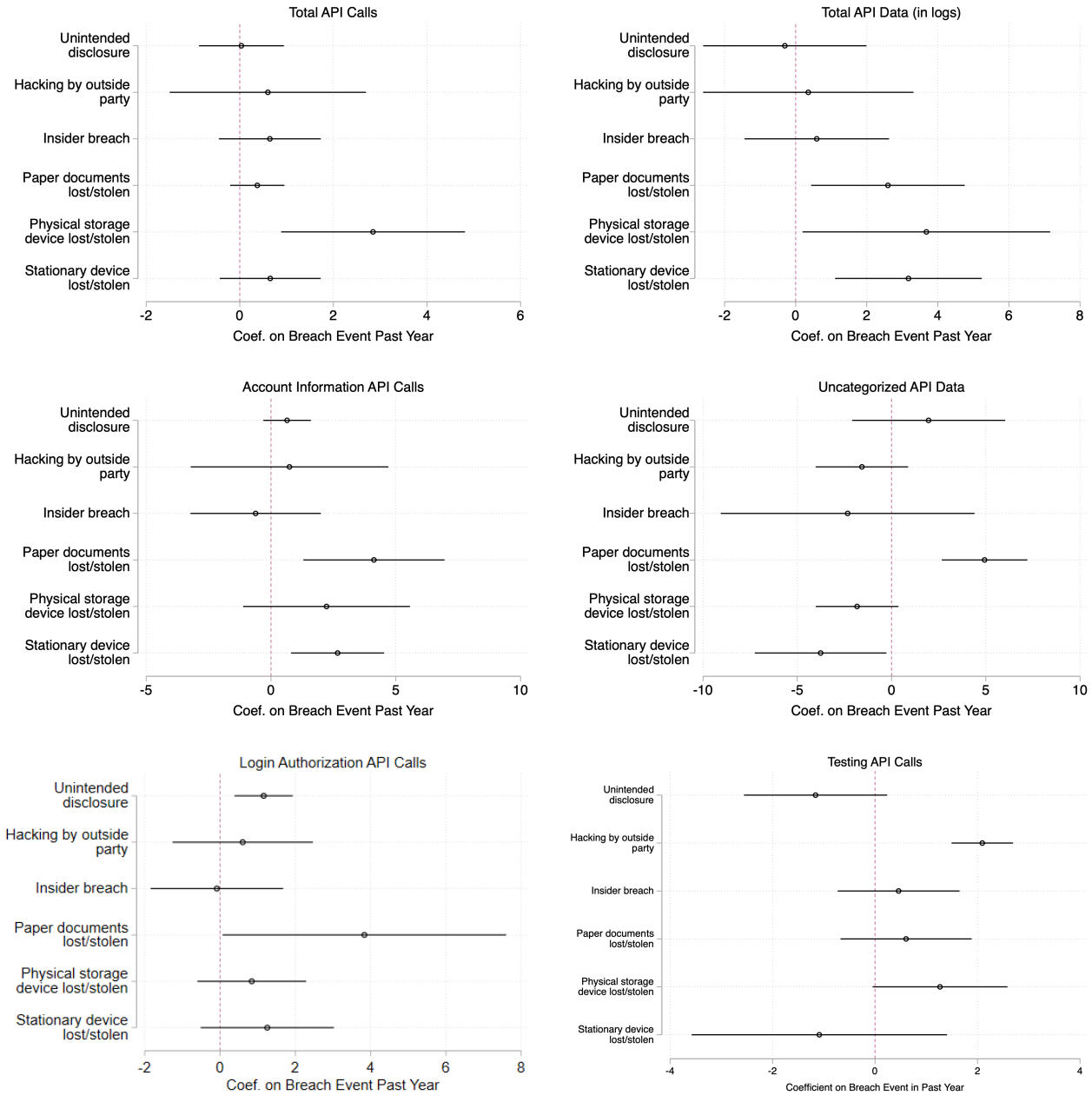
test calls [bottom right panel], while other sorts of data breaches, such as unintended disclosures do not. This confirms the earlier analysis and shows a logical response to a breach. 2) The theft of paper documents leads to a surprising increase in account API calls, login API calls, and total API calls [multiple panels]. This implies firms digitize their physical documents after a paper document theft and require authenticated access. 3) Following unintended disclosures and insider breaches, we do not detect changes in total API use [top right panel]. If anything we see a decrease in the use of security-focused APIs [bottom right panel]. This is consistent with firms reducing the number of individuals with security access after these types of breaches. 4) The loss of a physical storage device, such as a hard drive or thumb drive, leads to statistically significant increases in total API data [top right panel], account information API calls [middle left panel], and login authorization data [bottom left panel]. This seems intuitive as firms seek to secure their remaining systems against authorized access that stolen devices might provide. The overall picture provided by these estimates would illustrate a firm upgrading its security and automating more of its operational processes through APIs. These fine grained data provide a rich view of action and response inside a firm.[20]

VI. Conclusions and Managerial Implications

The paradox of openness is that APIs allow third parties to add value at the same time it may facilitate their misappropriation of value. Not all prospective partners are benign. To measure the tradeoff in exposure versus efficiency, we collected a unique data set combining firm-level financial performance, public announcements of data breaches, and exact measures of private API data flows through ten dozen firms. We tested four hypotheses, resulting in several findings. First, we estimate that API adoption led firms to have 19.1% higher market values in the year after adoption, and 20.1% higher values in subsequent years (Table IV). This increase, roughly one fifth the value of adopting firms, is a dramatic gain relative to the cost of introducing and maintaining APIs. Second, one mechanism of efficiency gain is a reduction in R&D expenditures. Controlling for firm and time fixed effects, firms that adopted APIs had

[20]We provide plots for all breaches and API types in Figures 19 and 20 in the Appendix.

Figure 16: Regression summary of the effect of a breach event on API use. Vertical axis shows breach type. Horizontal axis shows size of effect for coefficient on API usage of that type. Estimates follow equation 6 in the year after data breach announcement with 95% confidence intervals and clustered robust standard errors.



spent 25.3% more on R&D than average in the years prior to adoption, but spent no more than average two years after API adoption. Typically, they saved $34 million as firms substituted away from internal R&D. This is consistent with the theory of an inverted firm (Parker et al., 2017), in which production is moved from inside to outside and carried out by third parties rather than employees. Third, and unfortunately, in the three years after opening APIs, there

is a 1.19 percentage point increase in monthly risk of a data breach. This corresponds to .43 additional data breaches in expectation over the three year period.[21] Firms appear differentially skilled at handling cyberattacks. This is demonstrated by the dramatic increase in API call and data flow variance in the month before a hack is announced. Data transmissions spike at some firms and collapse at others, suggesting better monitoring leads some firms to limit their losses while others, unaware of pillaging, hemorrhage data at higher rates. Fourth, firms get better at using APIs over time. They see increasing market values and decreasing R&D expenditures. They respond to cyberattacks by deprecating non-standard APIs, reducing communications through compromised channels, and increasing testing. Yet, firms increase use of APIs, i.e. boost digital data flows, following the loss of physical media.

Although the "A" in "API" does not stand for aperture, that is a good metaphor for its function. Like the aperture of a camera, an API can be precisely calibrated to allow just the right amount of information through. It can also open too wide, creating unintentional disclosures. The delicate workings of digital apertures might also be unable to resist the fingers of hackers who seek to force them open wider. We provide evidence that API adopting firms see a greater risk from data breaches, yet this risk can be mediated by increasing the sophistication and security of a firm's APIs. Taking both these facts into consideration, how should executives think about using APIs at their firms?

The most important takeaway of our paper is that, while API adoption conveys real risks of exposure, the average firm in our sample did very well. A back of the envelope analysis suggests that shareholders face a .16% reduction in expected cumulative annual returns due to the direct effect of API induced hacks.[22] If opening APIs is associated with market capitalization gains on the order of 20% while risking a comparatively tiny negative effect from data breaches, the paradox resolves strongly in favor of opening APIs. For the average firm in our data set, API adoption was a very good decision on net.

Given the attractiveness of this gamble, why do more firms not use APIs? For those that do, what advice improves their effect? We offer three sets of ideas – caveats, practical advice,

[21] Calculated as the mean of a Binomial distribution with 1.19% breach probability and 36 independent trials.

[22] This is calculated as .43 additional data breaches per three year period, times a typical breach reduction in cumulative market return of 1.1% (Kamiya et al., 2018) divided by three years.

and principled advice.

One caveat is that these estimates are all *averages* and our estimates are all average treatment effects on the treated. If firms are good at forecasting the impact of technology use, it represents the average experience of the firms best suited to harness APIs. Standard economic theory suggests that the *marginal* firm should expect no net benefit from any change in strategy. Firms vary in terms of the upside they might achieve by adoption, and in terms of their capacity to handle cyberattacks.

As a further caveat, we cannot rule out the possibility that some underlying factor drives all three: API adoption, hack risk, and market value growth. For example, one possibility is that some firms are better strategically situated to utilize a platform strategy. This might directly cause them to grow more rapidly or engage in risky data processing activities. If these firms also tend to adopt APIs we might confound the impact of API adoption with the effect of being the type of firm that can adopt a platform strategy. In other words, API adoption might be the signal of a firm adopting a profitable platform strategy, but not the cause of profits itself. This is another reason our estimates of adoption effects might not be good predictors of the effect of a random new firm adopting an API strategy.

For practical advice, firms have several options to avoid the downside and realize the upside of APIs. We discovered that, following a data breach, firms increased the number of APIs and reduced data flows per API. This is consistent with breaking large multipurpose APIs into multiple smaller APIs, which improves control by partitioning access rights. Better modular design allows openness *with* control as distinct from openness *versus* control. More modular designs, together with increased monitoring and use of testing APIs and reducing use of non-transparently named APIs, should reduce exposure and the risk of falling off the "data cliff" shown in Figures 11 and 12. Consistent with Baldwin and Clark (2006), this can increase options value. Information security countermeasures can also be effective (Makridis and Dean, 2018), showing a $17-21\%$ return on investment. Further, if it is the case, as our data suggest, that firms substitute external R&D for internal R&D, firms should not take these third party investments for granted. Simply opening a developer portal is not sufficient. Instead, firms should publish access protocols, provide tools, and court developers at least as carefully as they

would court consumers (Ramji, 2017). If third parties are to add value, the process for adding it should be easy and those who do should be rewarded.

For principled advice, firms should avoid an agency problem in their compensation schemes. If executives disproportionately bear the costs of a data breach while shareholders disproportionately benefit from open APIs, incentives push executives to avoid APIs. As noted in the literature review, several data breaches tied to API flaws have become embarrassing national scandals. Cautious, inexpert, or socially conscious CIOs may fear the failures associated with not protecting consumers' private data. Given that corporate boards have reduced executive compensation over cyberattacks (Kamiya et al., 2018), executives charged with creating value will have different incentives than those charged for data breaches, even if API adoption does improve net market value.

A final observation concerns firms that provide APIs and the policies to which they should adhere. If cloud platforms are positioned to observe the data flows of firms whose APIs they manage, then, as seems likely, they might also be in a position to anticipate changes in the market value of those same firms. By analogy, Alibaba underwrites loans that outperform those of banks by using real time transactions data from its market to predict credit worthiness (Bloomberg, 2019). Advance private information might then be akin to insider information. Cloud platforms that observe real time API call data would seem to have an information advantage, as suggested in the introduction (Zhu and Liu, 2018), that could extend to investment, mergers and acquisitions, and other performance indicators. Our findings suggest this could indeed be the case and that such live data could be of rising importance not only to firms and shareholders but also to regulators seeking to ensure fairness of online markets.

References

Adner, R. and R. Kapoor (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal 31*(3), 306–333.

Baldwin, C. and K. Clark (2000). *Design rules: The power of modularity*, Volume 1. The MIT Press.

Baldwin, C. Y. and K. B. Clark (2006). The architecture of participation: Does code architecture mitigate free riding in the open source development model? *Management Science 52*(7), 1116–1127.

Baldwin, C. Y. and C. J. Woodard (2009). *Platforms Markets and Innovation*, Chapter The Architecture of Platforms: A Unified View, pp. 131–162. Cheltenham, U.K.: Edward Elgar Publishing.

Benzell, S. G. and E. Brynjolfsson (2019). Digital abundance and scarce genius: Implications for wages, interest rates, and growth.

Bloomberg (2019, July 28). Jack Ma's $290 billion loan machine is changing Chinese banking. *Bloomberg News*.

Boudreau, K. (2010). Open platform strategies and innovation: Granting access versus devolving control. *Management Science 56*(10), 1849–1872.

Boulton, C. (2017, February 8). Insurance spin-out rides api-driven strategy. *CIO Magazine*.

Brynjolfsson, E. and L. M. Hitt (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic PerspectivesVolume 14*(4), 23–48.

Brynjolfsson, E., D. Rock, and C. Syverson (2018). The productivity J-curve: How intangibles complement general purpose technologies. *NBER Working Paper*.

Burtch, G., S. Carnahan, and B. N. Greenwood (2018). Can you gig it? an empirical examination of the gig economy and entrepreneurial activity. *Management Science 64*(12), 5497–5520.

Chesbrough, H. and M. Van Alstyne (2015). Permissionless innovation. *Communications of the ACM 58*(8), 24–26.

Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Cusumano, M., A. Gawer, and D. Yoffie (2019). *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. Harper Business.

David, H., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017). The fall of the labor share and the rise of superstar firms. *NBER Working Paper*.

Evans, P. C. and R. C. Basole (2016). Revealing the api ecosystem and enterprise strategy via visual analytics. *Communications of the ACM 59*(2), 26–28.

Haranis, M. (2017, February 17). Cisco mounting massive open API offensive. *CRN*.

Harvey, P. (2017, March 20). AT&T adds APIs to help partners move more mobility products, services. *CRN*.

Henfridsson, O. and B. Bygstad (2013). The generative mechanisms of digital infrastructure evolution. *MIS quarterly*, 907–931.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Huckman, R., G. Pisano, D. Chen, and L. Kind (2012). Amazon web services. *Harvard Business School Case 9-609-048*.

Iyer, B. and J. C. Henderson (2010). Preparing for the future: Understanding the seven capabilities of cloud computing. *MIS Quarterly Executive 9*(2), 117–131.

Iyer, B. and M. Subramanian (2015, January). The strategic value of APIs.

Jacobson, D., G. Brail, and D. Woods (2011). *APIs: A strategy guide*. O'Reilly Media, Inc.

Joseph, S., V. Ludford, and B. McAllister (2016, September). Plugging in: Enabling the enterprise for the platform economy. Technical report, Gartner Research Board.

Kamiya, S., J.-K. Kang, J. Kim, A. Milidonis, and R. M. Stulz (2018). What is the impact of successful cyberattacks on target firms?

Khan, L. (2017). Amazon's antitrust paradox. *Yale Law Journal 126*(3), 710–805.

Lusch, R. F. and S. Nambisan (2015). Service innovation: A service-dominant logic perspective. *MIS quarterly 39*(1).

Makridis, C. and B. Dean (2018). Measuring the economic effects of data breaches on firm outcomes: Challenges and opportunities. *Available at SSRN 3044726*.

Malone, T. W., J. Yates, and R. I. Benjamin (1987). Electronic markets and electronic hierarchies. *Communications of the ACM 30*(6), 484–498.

Musser, J. (2013, Feb). API business models: 20 models in 20 minutes. `https://www.infoq.com/presentations/API-Business-Models/`. API Strategy Conference.

Newman, L. (2018, December 10). Google+ exposed data of 52.5 million users and will shut down. *Wired*.

Parker, G. and M. Van Alstyne (2017). Innovation, openness, and platform control. *Management Science 64*(7), 3015–3032.

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

Parker, G. G., M. W. Van Alstyne, and S. P. Choudary (2016). *Platform Revolution: How Networked Markets Are Transforming the Economy And How to Make Them Work for You*. New York: WW Norton & Company.

Programmable Web (2016). API Research.

Purvis, R. L., V. Sambamurthy, and R. W. Zmud (2001). The assimilation of knowledge platforms in organizations: An empirical investigation. *Organization science 12*(2), 117–135.

Ramji, S. (2017, July 14). Developer ecosystems in the age of platforms. `https://www.youtube.com/watch?v=Umahtle2aZE`. Keynote at the MIT Platform Summit.

Ransbotham, S. (2016). Open source code and the risk of attacks after vulnerability discovery. `http://programme.exordo.com/oui2016/delegates/presentation/140/`. Open and User Innovation Conference.

Schulze, E. (2019). How Google, Facebook, Amazon and Apple faced EU tech antitrust rules. `https://www.cnbc.com/2019/06/07/how-google-facebook-amazon-and-apple-faced-eu-tech-antitrust-rules.html`. CNBC.

Spanos, G. and L. Angelis (2016). The impact of information security events to the stock market: A systematic literature review. *Computers & Security 58*, 216–229.

Spring, T. (2018, August 24). T-mobile alerts 2.3 million customers of data breach tied to leaky API. *ThreatPost*.

Thierer, A. (2016). *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*. Mercatus Center at George Mason University.

Tilson, D., K. Lyytinen, and C. Sørensen (2010). Research commentarydigital infrastructures: The missing is research agenda. *Information Systems Research 21*(4), 748–759.

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

Yoo, Y., R. J. Boland, K. Lyytinen, and A. Majchrzak (2012). Organizing for innovation in the digitized world. *Organization Science 23*(15), 1398–1408.

Zeng, M. (2015). Three paradoxes of building platforms. *Communications of the ACM 58*(2), 27–29.

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

VII. Online Appendix

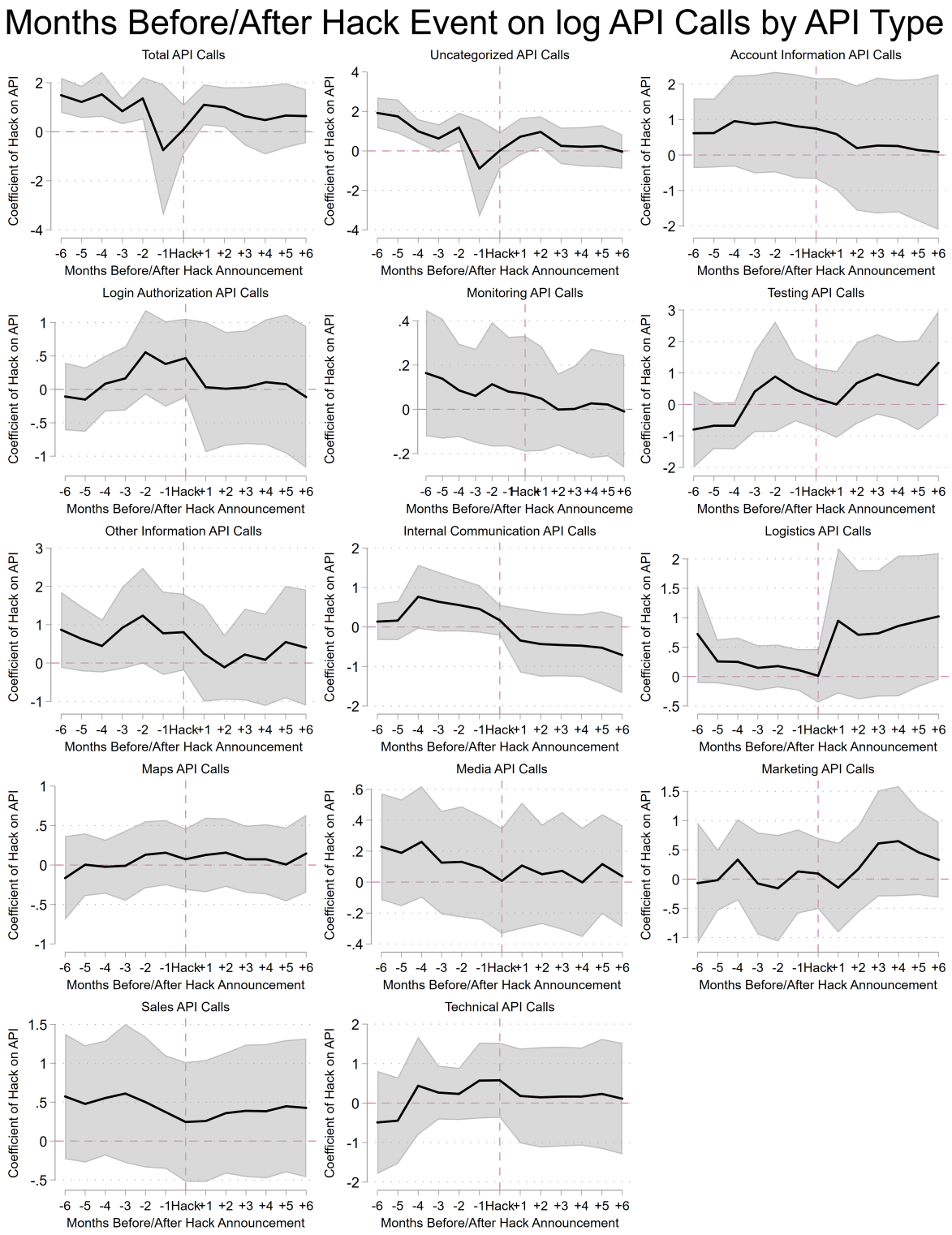


Figure 17: Hack Announcement and Immediate API Calls

Table X: Estimates of API adoption on market value and R&D using quarterly intervals

	log Market Value	log R&D
6+ quarters until API adoption	-0.0811	0.231***
	(0.111)	(0.0551)
5 quarters until API adoption	0.0254	0.152**
	(0.0879)	(0.0477)
4 quarters until API adoption	0.0405	0.172***
	(0.0879)	(0.0486)
3 quarters until API adoption	0.0674	0.186***
	(0.0855)	(0.0487)
2 quarters until API adoption	0.0879	0.139**
	(0.0849)	(0.0497)
1 quarters until API adoption	0.0833	0.122**
	(0.0802)	(0.0399)
Adoption Quarter	0.101	0.105**
	(0.0799)	(0.0360)
1 quarters since API adoption	0.0889	0.109**
	(0.0756)	(0.0367)
2 quarters since API adoption	0.120	0.0899+
	(0.0739)	(0.0496)
3 quarters since API adoption	0.140*	0.0619
	(0.0662)	(0.0411)
4 quarters since API adoption	0.155*	0.0736
	(0.0674)	(0.0520)
5 quarters since API adoption	0.156*	0.0589
	(0.0685)	(0.0415)
6+ quarters since API adoption	0.208***	0.00210
	(0.0564)	(0.0345)
Constant	5.318***	1.319***
	(0.0116)	(0.00941)
Firm FEs	Yes	Yes
Quarter FEs	Yes	Yes
N	316738	141484
Firms	12849	6829
API Adopters	55	55
Balanced	No	No
Additional Controls	No	No
Only Firms w/ Dev Portals	No	No

Notes: Std errs in parenthesis. SEs clustered at the firm level. Additional controls include firm level net goodwill, total revenue, long term debt, and operating expenses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

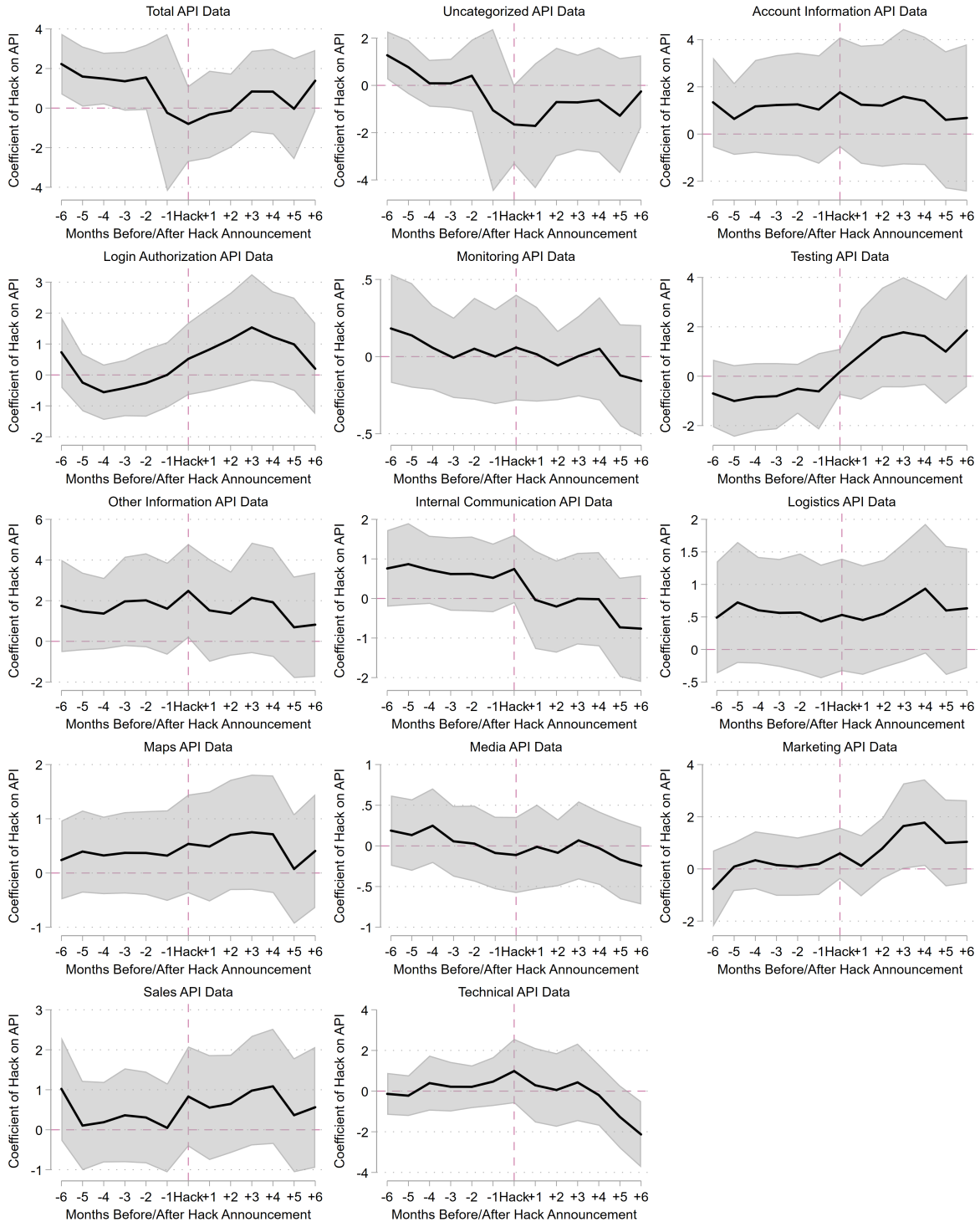


Figure 18: Hack Announcement and Immediate Call Trends

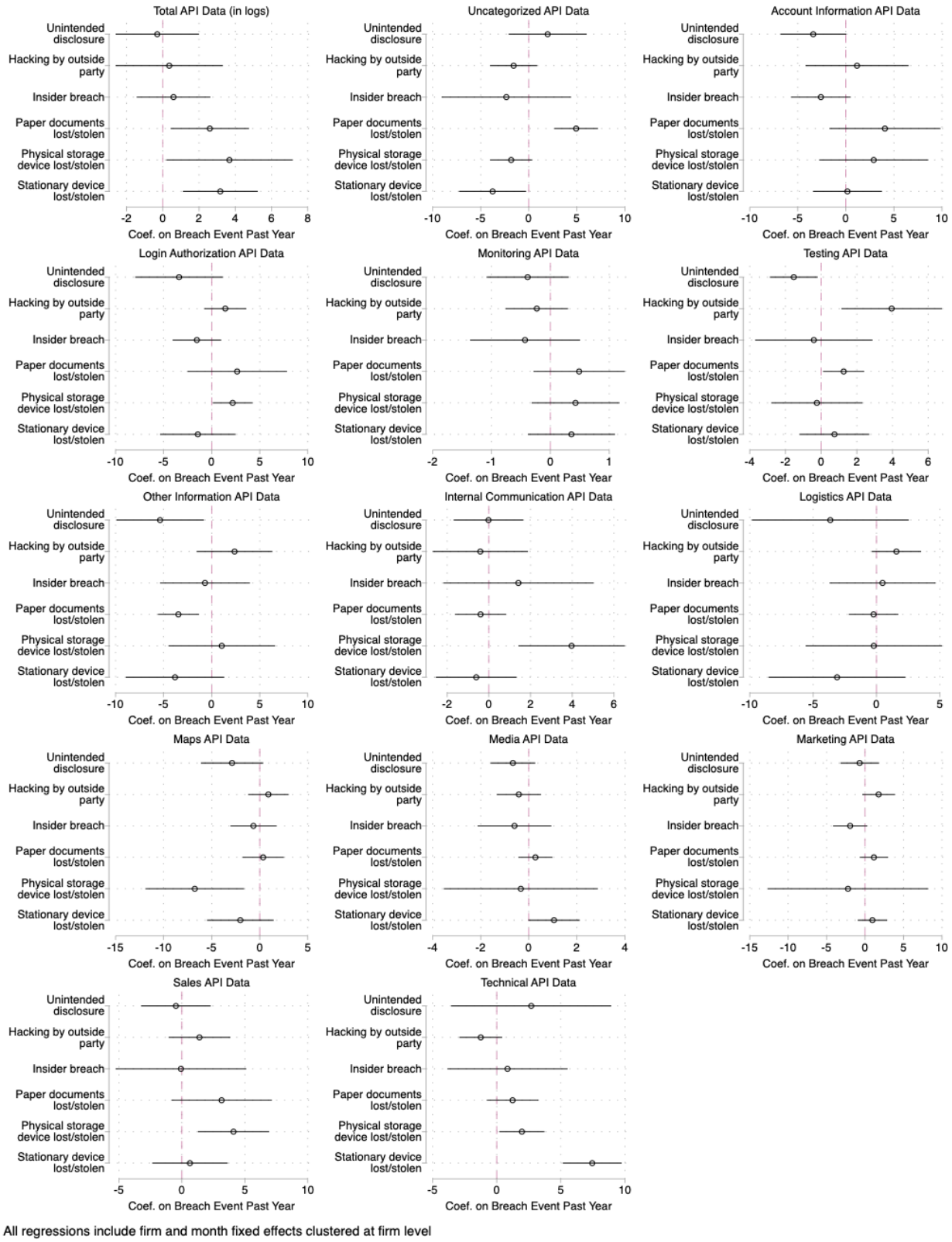


Figure 19: Regression Summary Coefficient Plot: Impact of Type of Security Breach on API Use Intensity (log API Data Volume) by Classification of API

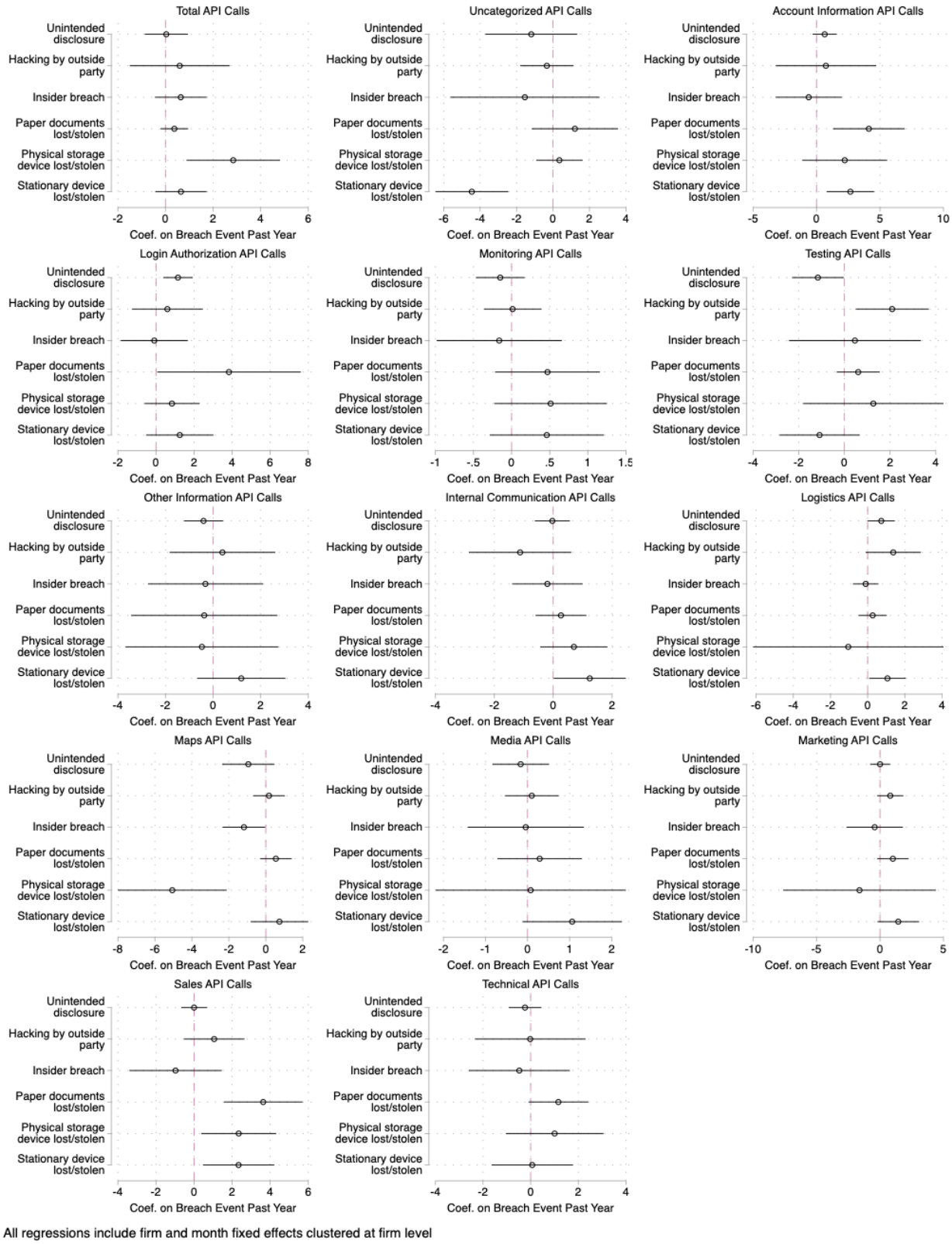


Figure 20: Regression Summary Coefficient Plot: Impact of Type of Security Breach on API Extensive Margin Intensity (log API Data Calls) by Classification of API

Table XI: Impact of Type of Security Breach on API Use Intensity (log Data Volume) by Function of API

Dependent variable: data transfer by API type	Total API Data	Un-categorized	Account Info	Login Autho-rization	Monitoring	Testing	Other Info	Internal Comm	Logistics	Maps	Media	Marketing	Sales	Technical
Type of Breach:														
Unintended disclosure	-0.304	1.966	-3.387^{+}	-3.388	-0.387	-1.531*	-5.367*	-0.0159	-3.643	-2.876^{+}	-0.660	-0.705	-0.472	2.694
	(1.160)	(2.051)	(1.731)	(2.308)	(0.352)	(0.676)	(2.302)	(0.847)	(3.122)	(1.637)	(0.470)	(1.261)	(1.396)	(3.156)
Hacking/malware by outside party	0.356	-1.571	1.166	1.415	-0.231	3.952**	2.382	-0.412	1.572	0.904	-0.412	1.768	1.398	-1.245
	(1.495)	(1.238)	(2.708)	(1.107)	(0.268)	(1.423)	(1.989)	(1.154)	(0.983)	(1.063)	(0.465)	(1.070)	(1.239)	(0.846)
Insider breach	0.594	-2.329	-2.595	-1.539	-0.431	-0.407	-0.698	1.418	0.481	-0.652	-0.595	-1.935^{+}	-0.0740	0.843
	(1.027)	(3.401)	(1.566)	(1.272)	(0.472)	(1.661)	(2.362)	(1.819)	(2.114)	(1.213)	(0.776)	(1.129)	(2.622)	(2.370)
Paper documents lost, discarded or stolen	2.596*	4.936***	4.070	2.651	0.489	1.265*	-3.462**	-0.398	-0.233	0.357	0.278	1.153	3.155	1.239
	(1.092)	(1.147)	(2.920)	(2.622)	(0.392)	(0.580)	(1.082)	(0.619)	(0.988)	(1.093)	(0.359)	(0.931)	(2.017)	(1.022)
Physical storage device (hard drive, etc) lost or stolen	3.678*	-1.828	2.902	2.188*	0.427	-0.238	1.048	3.969**	-0.215	-6.759*	-0.329	-2.219	4.105**	1.968*
	(1.760)	(1.108)	(2.867)	(1.040)	(0.377)	(1.291)	(2.802)	(1.290)	(2.716)	(2.594)	(1.620)	(5.273)	(1.430)	(0.888)
Stationary computer lost or stolen	3.174**	-3.762*	0.169	-1.439	0.357	0.748	-3.809	-0.605	-3.107	-2.018	1.053^{+}	0.953	0.642	7.456***
	(1.043)	(1.764)	(1.814)	(1.986)	(0.374)	(0.986)	(2.585)	(0.981)	(2.728)	(1.752)	(0.540)	(0.966)	(1.513)	(1.159)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.768	0.628	0.177	0.155	0.0615	0.185	0.221	0.120	0.106	0.149	0.0811	0.190	0.205	0.190
Obs	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535
Firms	123	123	123	123	123	123	123	123	123	123	123	123	123	123

Notes: Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable of each column represent log volume of data transferred through APIs classified by API purpose. Each independent variable is dummy coded as 1 if the firm experienced a data breach of a given type within the past year.

Table XII: Impact of Type of Security Breach on API Use Intensity (log API Calls) by Function of API

Dependent variable: API calls by API function	Total API Calls	Un-categorized	Account Info	Login Autho-rization	Monitoring	Testing	Other Info	Internal Comm	Logistics	Maps	Media	Marketing	Sales	Technical
Type of Breach:														
Unintended disclosure	0.0352	-1.194	0.645	1.159**	-0.149	-1.161*	-0.400	-0.0277	0.734*	-0.946	-0.164	0.00590	0.00497	-0.236
	(0.458)	(1.274)	(0.481)	(0.391)	(0.161)	(0.571)	(0.413)	(0.301)	(0.369)	(0.715)	(0.339)	(0.396)	(0.339)	(0.345)
Hacking/malware by outside party	0.600	-0.337	0.743	0.602	0.0140	2.093*	0.393	-1.123	1.379+	0.172	0.101	0.807	1.058	-0.0181
	(1.060)	(0.738)	(2.000)	(0.941)	(0.191)	(0.809)	(1.121)	(0.878)	(0.749)	(0.435)	(0.324)	(0.523)	(0.804)	(1.175)
Insider breach	0.643	-1.543	-0.615	-0.0857	-0.162	0.460	-0.319	-0.197	-0.105	-1.184*	-0.0431	-0.418	-0.971	-0.477
	(0.550)	(2.069)	(1.320)	(0.889)	(0.415)	(1.458)	(1.221)	(0.606)	(0.347)	(0.583)	(0.698)	(1.120)	(1.226)	(1.072)
Paper documents lost, discarded or stolen	0.375	1.203	4.125**	3.834*	0.471	0.606	-0.371	0.264	0.272	0.545	0.289	1.016	3.635***	1.165+
	(0.293)	(1.196)	(1.429)	(1.903)	(0.348)	(0.476)	(1.552)	(0.437)	(0.383)	(0.433)	(0.508)	(0.627)	(1.043)	(0.634)
Physical storage device (hard drive, etc) lost or stolen	2.848**	0.354	2.227	0.842	0.512	1.267	-0.468	0.703	-1.044	-5.058***	0.0737	-1.613	2.346*	1.009
	(0.992)	(0.644)	(1.688)	(0.729)	(0.374)	(1.551)	(1.624)	(0.577)	(2.588)	(1.488)	(1.143)	(3.040)	(0.994)	(1.036)
Stationary computer lost or stolen	0.651	-4.454***	2.669**	1.254	0.462	-1.087	1.185	1.243*	1.066*	0.737	1.061+	1.440+	2.341*	0.0668
	(0.545)	(1.010)	(0.942)	(0.894)	(0.379)	(0.886)	(0.938)	(0.619)	(0.496)	(0.788)	(0.598)	(0.828)	(0.946)	(0.864)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.265	0.126	0.165	0.113	0.0931	0.172	0.151	0.112	0.0759	0.145	0.104	0.168	0.229	0.156
Obs	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535	2535
Firms	123	123	123	123	123	123	123	123	123	123	123	123	123	123

Notes: Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable of each column represent log volume of calls to an APIs classified by API purpose. Each independent variable is dummy coded as 1 if the firm experienced a data breach of a given type within the past year.

A. API Orientations and Functions

Using our proprietary dataset from an API tool provision company, we categorized APIs by their orientation and function. The orientation of API indicates whether its primary purpose is for interactions with consumers, internal systems or employees, or other businesses.

In addition to an API's orientation, we also categorize APIs by their function. We sorted APIs into the following functions:

- Account Information: APIs related to storing, retrieving and displaying users profiles
- Internal Communication: APIs for internal communication between employees
- Login/Authorization: APIs authenticating users and allowing information to be securely shared with other platforms
- Logistics/Inventory: APIs related to recording, managing and optimizing logistical items and inventory flow such as order delivery
- Maps/Locations: APIs dedicated to maps and GPS platforms, often Google Maps.
- Marketing/Customer Insights/Analytics: APIs related to storing and/or analyzing customer behavior or advertising information
- Media: APIs related to accessing, displaying or linking news or social media content
- Monitoring/Data Traffic Management: APIs related to collecting and managing data traffic
- Other: Identified APIs storing and providing information but unrelated to standard categories
- Sales: APIs related to consumer purchases, especially online shopping
- Test: Any API named a variation on 'test' as well as any other API used for conducting tests of the platform performance
- Technical: APIs performing technical internal function task unrelated to the aforementioned categories
- Uncategorized: APIs whose function could not be discerned from the name, the company developer portal, or Internet search

Many APIs have names which directly point to their functions, such as sales or login APIs. To determine the function of APIs with unclear or technical names, we did additional research. Internet search of technical API names often revealed their function. There was also often information on a firm's developer portal.

After classifying hundreds of APIs manually, we were able to identify consistent relationships between API names and corresponding functions. Using these relationships, we were able to

identify and use certain keywords to partially automate API categorization. All automatic categorizations were double checked by hand.

Occasionally, even after additional research, how an API should be classified remained ambiguous. For example, APIs such as "Pingdom" performed tasks falling in both the Monitoring and Test categories. Similarly, APIs classified as Marketing or Sales could often arguably be placed in the other category. We used our best judgment in the classification of these ambiguous cases.