

Nowcasting Firm Performance and Data Breaches Using API Data

Seth Gordon Benzell, Jonathan Hersh, Guillermo Lagarda & Marshall
Van Alstyne

MIT IDE, Chapman University, IADB, Boston University
Funding by Accenture, Apigee & Mulesoft is gratefully acknowledged

October 29, 2018

Introduction

- Modern firms face trade-off between information openness and security
- APIs (application program interfaces) can greatly increase data fluidity within their firm and to other entities

Introduction

- Modern firms face trade-off between information openness and security
- APIs (application program interfaces) can greatly increase data fluidity within their firm and to other entities
- How valuable is adopting an open API strategy?
- What is the information content of those flows?
- Are there unexpected costs?

Introduction

- We use a novel firm-level panel of API adoption and API data flows by **orientation** (B2B, etc) and **type** (tech, marketing)
- 124 firms between 2007-2016 matched to Compustat quarterly/monthly firm outcomes

Introduction

- We use a novel firm-level panel of API adoption and API data flows by **orientation** (B2B, etc) and **type** (tech, marketing)
- 124 firms between 2007-2016 matched to Compustat quarterly/monthly firm outcomes

We investigate:

- 1 Benefits to adopting APIs in terms of firm market value and other outcomes (+)
- 2 Whether API data can nowcast firm performance (mixed)
- 3 Impact of API adoption on probability of cyberattacks and information disclosures (APIs → ↑ hacks)

Introduction

- We use a novel firm-level panel of API adoption and API data flows by **orientation** (B2B, etc) and **type** (tech, marketing)
- 124 firms between 2007-2016 matched to Compustat quarterly/monthly firm outcomes

We investigate:

- 1 Benefits to adopting APIs in terms of firm market value and other outcomes (+)
- 2 Whether API data can nowcast firm performance (mixed)
- 3 Impact of API adoption on probability of cyberattacks and information disclosures (APIs → ↑ hacks)

This rationalizes the puzzle of why all firms have not adopted an API strategy

Literature Review

- API adoption has a very positive financial impact

(Benzell, Lagarda, Van Alstyne, 2017)

- Communication is one of the key tasks of firms

(Coase, 1937; Argote, McEvily, Reagans 2003; Galbraith 2007; Aral, Brynjolfsson, Van Alstyne 2012)

- Digital flows can be used to nowcast current events

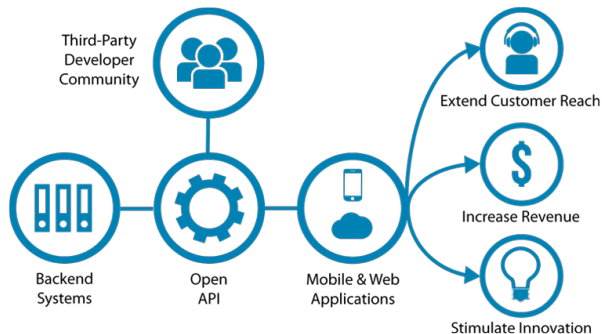
(Choi and Varian, 2012)

- Hacks lead to average loss of \$439 million per attack

(Kamiya et. al. 2012)

APIs

- Application Programming Interfaces (APIs) are software contracts that allow one piece of code to access the functions of another.
- They are building blocks of digital ecosystems, enhance modularity, and facilitate metering.



Costs & Benefits of APIs

Why use APIs?

- Modularity, reuse ↑
(Verizon new phones)
- Agility, efficiency ↑, Costs ↓
(Cleveland Clinic EMR)
- Sales channels ↑, 3rd party products ↑
(Walgreens)
- Market Capitalization ↑
(Amazon)

Costs & Benefits of APIs

Why use APIs?

- Modularity, reuse ↑
(Verizon new phones)
- Agility, efficiency ↑, Costs ↓
(Cleveland Clinic EMR)
- Sales channels ↑, 3rd party products ↑
(Walgreens)
- Market Capitalization ↑
(Amazon)

Why *not* use APIs?

- Failure risk ↑ (no APIs on pacemakers!)
- Hack risk ↑ (TJX, Experion)
- Data loss ↑ (Netflix)
- Support Costs ↑ (Netflix)
- Competing Apps ↑
(Google maps, Twitter)

Costs & Benefits of APIs

Why use APIs?

- Modularity, reuse ↑
(Verizon new phones)
- Agility, efficiency ↑, Costs ↓
(Cleveland Clinic EMR)
- Sales channels ↑, 3rd party products ↑
(Walgreens)
- Market Capitalization ↑
(Amazon)

Why *not* use APIs?

- Failure risk ↑ (no APIs on pacemakers!)
- Hack risk ↑ (TJX, Experion)
- Data loss ↑ (Netflix)
- Support Costs ↑ (Netflix)
- Competing Apps ↑
(Google maps, Twitter)

Data Sources

- Novel proprietary API use data. Categorized by API Function (tech, marketing, etc) and Orientation (B2B, etc)
- Matched to Compustat monthly and quarterly
 - 124 firms
 - 42 Information
 - 12 Finance
 - 25 manufacturing (31-33)
 - 19 Retail -Trade
 - 2535 firm-months of API use

Data Sources

- Novel proprietary API use data. Categorized by API Function (tech, marketing, etc) and Orientation (B2B, etc)
- Matched to Compustat monthly and quarterly
 - 124 firms
 - 42 Information
 - 12 Finance
 - 25 manufacturing (31-33)
 - 19 Retail -Trade
 - 2535 firm-months of API use
- Firm Developer Portal Use
- Merged to PRC Data Breach Records (as in Kamiya et. al. 2012)
 - Firm discloses information breach
 - How many records breached



APIs

The Walgreens Developer Program allows you access to tap into the large and growing number of Walgreens stores, products, and more! Including the ability to print photos for same-day pickup, refill prescriptions, and clip coupons.

Photo Prints API

Photo Prints API allows any create/edit/share/store photo application developers the ability to print photos to any of the 8,000+ Walgreens and Duane Reade stores for Same-Day pickup. With every photo order that is successfully placed and picked up at a store, the developer can earn a revenue share commission.

[HTML Checkout Documentation](#)

[JSON Checkout Documentation](#)

[Prescription Refill/Transfer APIs](#)

Data Description

| API Function | Total Data Flow (Gigabytes) | Total Calls (Billions) | Data SD Across Firm-Months | Data SD Across APIs | Call SD Across Firm-Months | Call SD Across APIs |
|---------------------|-----------------------------|------------------------|----------------------------|---------------------|----------------------------|---------------------|
| Account Info | 51830 | 14.43 | 218 | 15.180 | 0.04 | 0.001641 |
| Other Info | 582100 | 75.22 | 1502 | 135.600 | 0.20 | 0.018235 |
| Internal Commun. | 21820 | 3.36 | 80 | 4.144 | 0.01 | 0.000535 |
| Login Auth. | 345400 | 21.75 | 2239 | 21.870 | 0.06 | 0.001631 |
| Logistics | 350500 | 24.24 | 1307 | 49.990 | 0.07 | 0.002236 |
| Maps | 266100 | 9.41 | 1291 | 78.500 | 0.03 | 0.001204 |
| Media | 5327 | 2.74 | 23 | 0.001 | 0.01 | 2.84E-07 |
| Marketing + Loyalty | 122500 | 17.54 | 401 | 16.200 | 0.04 | 0.001074 |
| Data Monitoring | 1384 | 0.58 | 9 | 0.263 | 0.01 | 8.5E-05 |
| Sales | 59440 | 11.20 | 283 | 4.480 | 0.03 | 0.000823 |
| Technical | 76120 | 68.58 | 264 | 6.009 | 0.25 | 0.000456 |
| Testing | 2249 | 0.52 | 13 | 0.105 | 0.01 | 6.47E-06 |
| Uncategorized | 1851000 | 143.40 | 9066 | 84.510 | 0.37 | 0.005875 |

Table 1: Total API flows observed for the 124 firms in the data. Firms in data starting first month with data flow. December 2012 through October 2016.

Table 2: Financial Variables in Thousands of Dollars

| Financial Variable | Average | SD |
|-------------------------|----------|----------|
| Market Value | 36629.4 | 51298.2 |
| Assets Total | 1220208 | 16570278 |
| Inventories Total | 95091.68 | 1282941 |
| Cash | 2626.823 | 4012.162 |
| Goodwill net | 12123.4 | 61986.08 |
| Pretax Income | 35495.09 | 523973.1 |
| Revenue Total | 237650.4 | 3398809 |
| Cost of Goods Sold | 121803.4 | 1710821 |
| Operating Expense Total | 181865.6 | 2573602 |

API Data Description

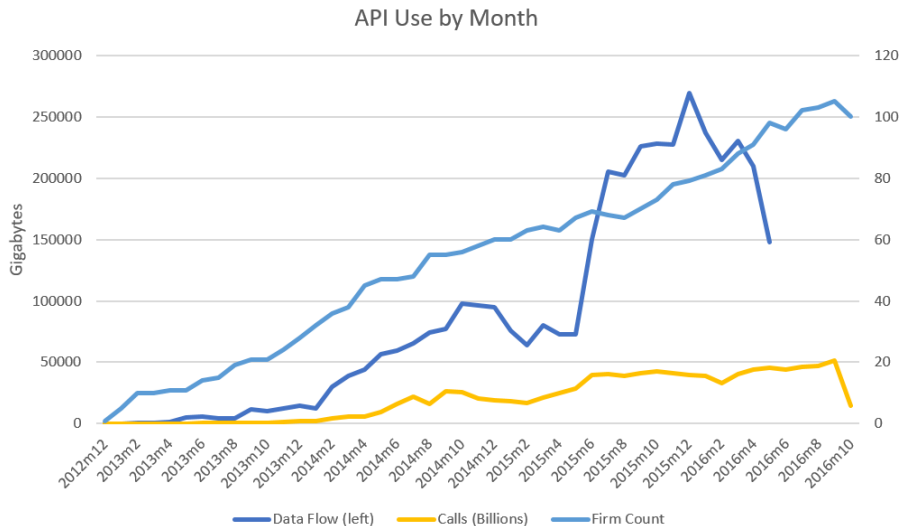
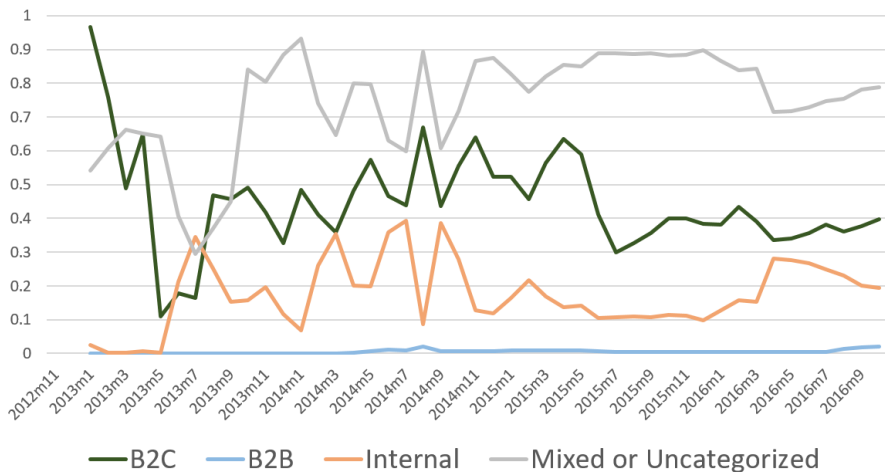


Figure 2: API Use by Month

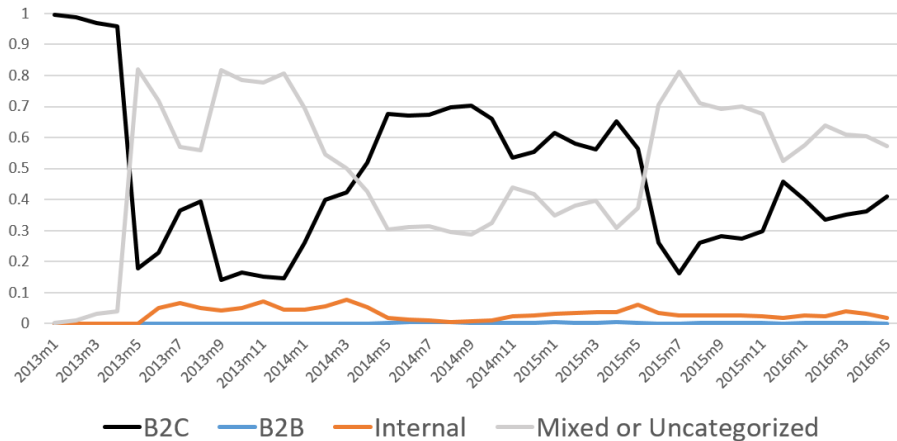
API Data Description

Share of API Calls by Orientation



API Data Description

Data Flow Share by Orientation



Nowcasting Using Random Forest

| | AR1 + API Vars | | AR1 Only | |
|---------------|----------------|-----------|------------|------------|
| | Training | Test | Training | Test |
| | RMSE | RMSE | RMSE | RMSE |
| Assets | 0.1741229 | 0.1639718 | 0.09303354 | 0.06605369 |
| Revenue | 0.2043672 | 0.2085332 | 0.15682995 | 0.15052152 |
| Income | 0.3761504 | 0.6968862 | 0.67582447 | 0.93963776 |
| Inventories | 0.2073757 | 0.2201898 | 0.12082855 | 0.14681152 |
| Cost of Goods | 0.1941133 | 0.2409644 | 0.1799582 | 0.17643629 |
| Goodwill | 0.1786985 | 0.3201358 | 0.27340076 | 0.18481797 |
| Cash | 0.1867028 | 0.4307641 | 0.36679337 | 0.461663 |
| Expenses | 0.1950493 | 0.7786918 | 0.15522271 | 0.96988574 |

- Separate random forest models fit for each outcome variable (Y_i).
500 trees per forest. X_{it} all data/flow/calls for all APIs by type.
Compared to AR(1): $y_{it} = \beta * y_{it-1} + \epsilon_{it}$

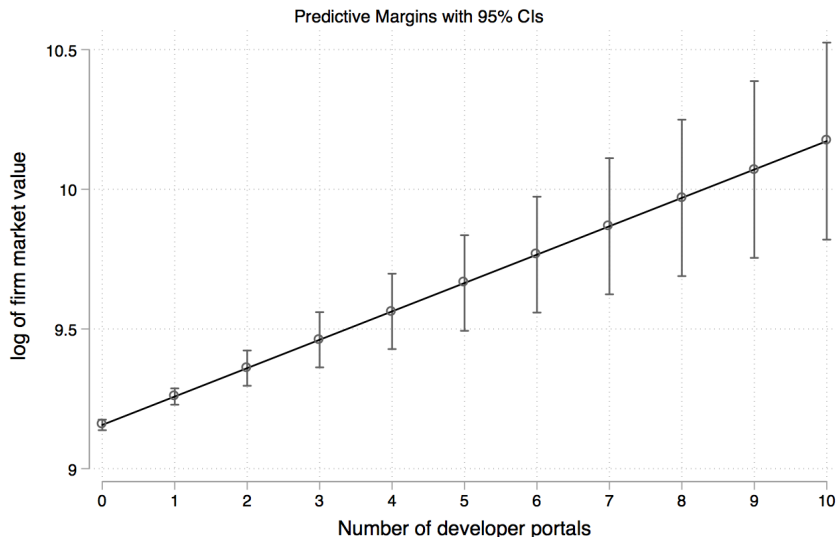
Impact of API Adoption on Firm Market Value

- Previous research (Benzell, Lagarda, Van Alstyne) has shown significant, large positive effect of API adoption on firm outcomes.
- Identification strategy is event study around date of first API use
- $y_{it} = \beta * API_Post4Years_{it} + \psi_i + \gamma_t + \epsilon_{it}$

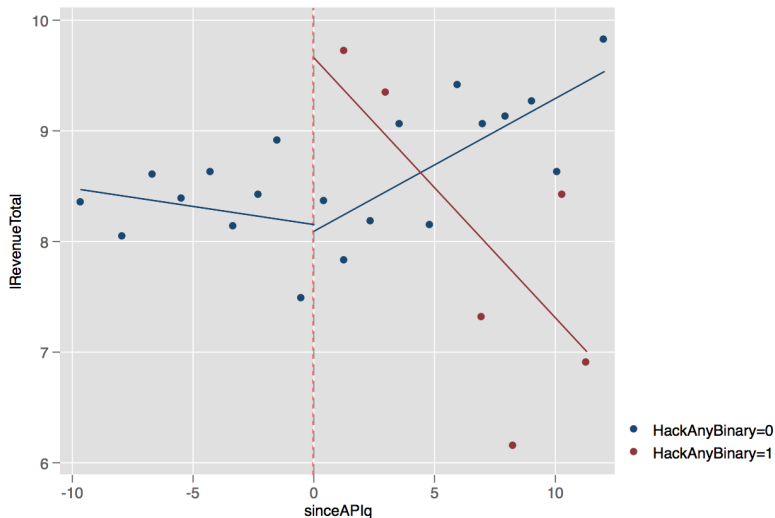
| | log Market Value | log Market Value |
|--|----------------------|----------------------|
| API Adoption | 0.125** (3.00) | -0.369*** (-3.71) |
| API and Number of Developer Portals | | 0.102*** (5.47) |
| Constant | 8.871*** (189.89) | 8.869*** (191.15) |
| N | 2212 | 2212 |

t statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Benefit of APIs Increasing with API Intensity (Number of Open APIs)



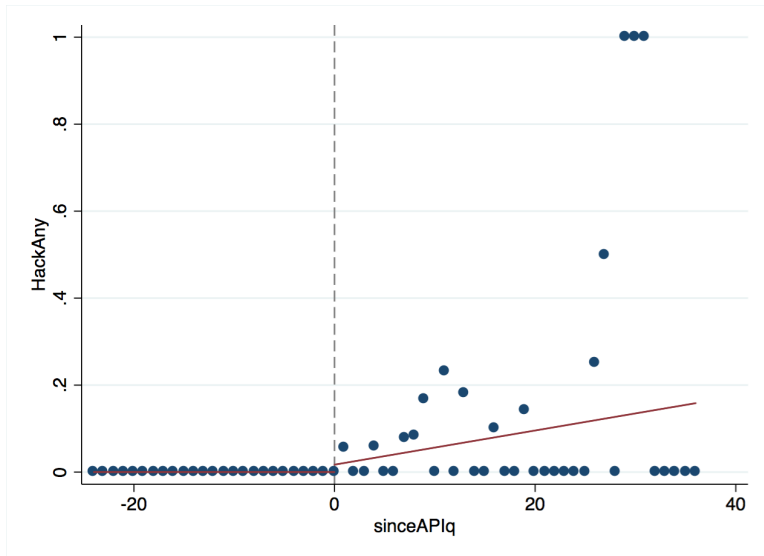
APIs are Great. What Could Go Wrong?



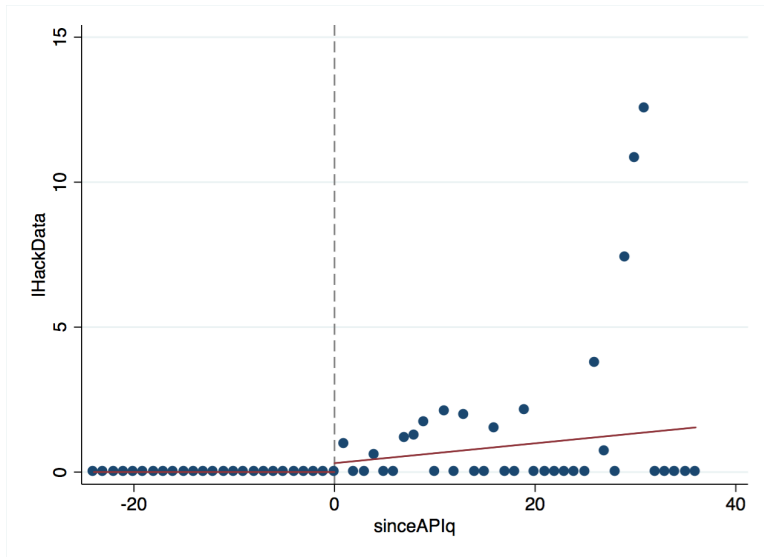
This Suggests Two Lines of Inquiry

- ① What happens to API data and calls following data-hack events?
- ② Which API strategies are more/less robust to data-hack events?

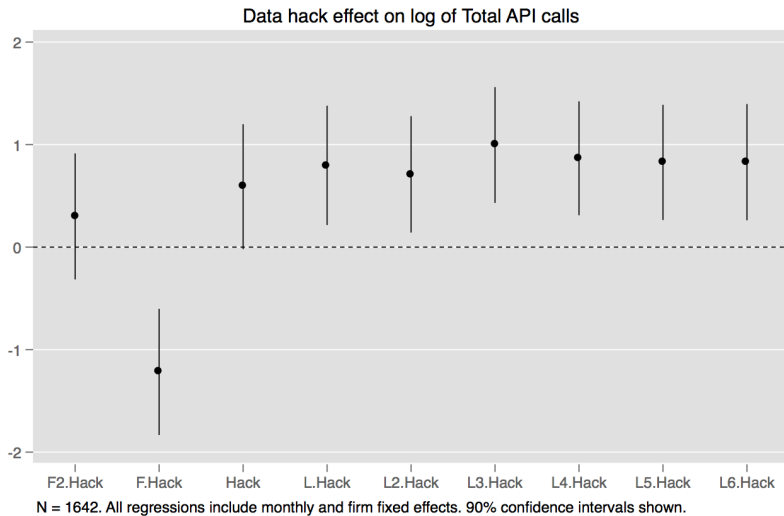
Probability of Data Breach Event Around API Adoption



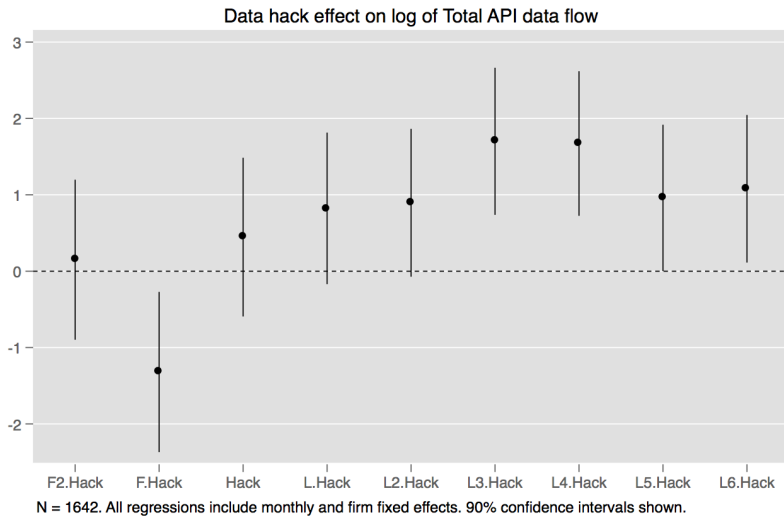
Number of Records Breached Around API Adoption



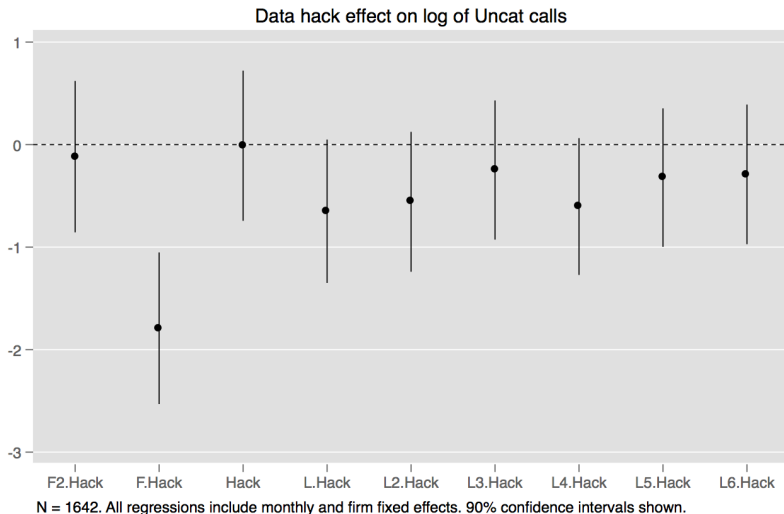
Total API Calls Pre- and Post- Data Breach Event



Total API Calls Pre- and Post- Data Breach Event



Total Calls Through Uncategorized API Calls Pre- and Post- Data Breach Events



| | Data Breach | Data Breach | Data Breach | Data Breach | Data Breach |
|--------------------------|----------------------|------------------------|----------------------|----------------------|-------------------------|
| Post API | 0.00965** (2.60) | 0.00654 (0.43) | 0.00781 (0.52) | 0.00764 (0.50) | 0.00756 (0.50) |
| Total Calls | | 3.88e-11* (2.50) | | | 5.43e-11** (3.03) |
| Total Calls ² | | -9.98e-21** (-3.00) | | | -1.49e-20*** (-3.63) |
| Num of APIs | | | 0.00000684 (0.03) | | -0.0000282 (-0.13) |
| Num of APIs ² | | | -2.16e-08 (-0.03) | | -6.05e-08 (-0.08) |
| Total Data | | | | 1.78e-16 (0.17) | -1.43e-15 (-1.23) |
| Total Data ² | | | | -1.61e-30 (-0.18) | 1.86e-29+ (1.75) |
| Constant | -2.445*** (-5.97) | 16.90** (2.88) | 13.27* (2.10) | 13.22* (2.35) | 17.06** (2.64) |
| N | 15528 | 2535 | 2535 | 2535 | 2535 |

To API or Not API: Hack is the Question

- Using APIs boosts market value by 12-13 percent
- But: firms who adopt an API strategy experience 0.12 more data breaches per year
- How risk loving would you have to be to accept this deal?

To API or Not API: Hack is the Question

- Using APIs boosts market value by 12-13 percent
- But: firms who adopt an API strategy experience 0.12 more data breaches per year
- How risk loving would you have to be to accept this deal?
- Data breach propensity increasing in calls, but decreasing in data
- Implication: send more data per calls
- Firms reduce their API flows in the month before announcement of a hack
- This is especially true of uncategorized API types

Future Work

- Investigate further the mechanism relating API use to data breaches
- Use model of Network Effects to explore optimal locus of work
- Use Agency and Incomplete Contract Theory to link API Functions to Firm Organization