

Website Changes and User Behavior

Using Panjiva Data to Examine Code Changes

John J. Wang

14.27 Final Presentation

November 30, 2012

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

The Silicon Valley Mindset

- "Move fast and break things." – Facebook
- "The only constant is change itself." – Heraclitus
- "Pick a movement, pick a revolution, and join it." – Jack Dorsey

1 Introduction

- Silicon Valley Mindset
- **What We Know**
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

Background and Previous Research

- Academia has little to say on code changes and user behavior
- Most of the data is hidden away in large tech companies
- Although these companies probably run experiments, results aren't necessarily made public

1 Introduction

- Silicon Valley Mindset
- What We Know
- **Questions to Ask**
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

Lingering Questions

- Do users tend to respond favorably to website changes?
- How do users react to different types of change?
- How do different characteristics of users affect their reactions to change?

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- **Panjiva Dataset**

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

Panjiva, Inc.

- <http://www.panjiva.com>
- Acts as a medium for buyers and suppliers of manufactured goods
- Parses government import and export data for unbiased shipment information on suppliers
- Example: Home Depot finding a wrench factory

Summary Statistics

Table: Panjiva Overview

Total Users	121,653
Subscribing Users	2,985
Monthly Site Visits	903,426
Monthly Unique Visitors	762,723
Average Pages per Visit	1.99
Average Visit Duration	1 min 18 sec

Event and Activity Logs

- Event Logs contain records of all user actions 124 million records
- Activity Logs contain records of actions of users who are registered with Panjiva 13 million records

Commit Statistics

Table: Overall Commit Statistics - 11/25/2012

Active Days (at least 1 commit)	1,983
Total Current Files	20,901
Total Lines of Code	1,313,235
Total Lines of Code Added	3,989,295
Total Lines of Code Removed	2,676,060
Total Commits	29,924
Total Authors/Developers	33

Commit Frequency by Time of Day

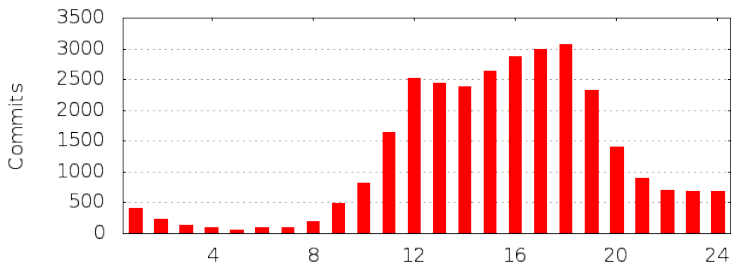


Figure: Number of Commits by Hour of the Day

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

Specification

$$y_t^i = c_0 + \vec{\gamma}^T \vec{M}_t + \vec{\beta}^T \vec{\chi}_t + \epsilon_t \quad (1)$$

- t indexes day
- y_t^i corresponds to i th metric of user activity on day t
- \vec{M}_t corresponds to a vector of covariates that represent changes in the code
- $\vec{\chi}_t$ is a vector of controls

Effect of Commits on User Activity

	(1) activitylogcount	(2) eventlogcount
fileschanged percentile	-5530.7 (-1.91)	-60589.4* (-2.26)
insertions percentile	4868.8* (2.12)	47053.7* (2.22)
deletions percentile	2778.9 (1.29)	29970.6 (1.50)
weekend	-14708.0*** (-16.48)	-79769.8*** (-9.65)
_cons	22396.6*** (25.05)	224482.8*** (27.12)
<i>N</i>	474	475

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Effect of Commits on User Activity

	(1) avguseractivity	(2) avguserevents
fileschanged	-10.40*	-148.6
percentile	(-2.01)	(-1.52)
insertions	2.817	79.92
percentile	(0.69)	(1.03)
deletions	4.670	49.47
percentile	(1.21)	(0.68)
weekend	-0.297 (-0.19)	257.0*** (8.49)
_cons	25.00*** (15.58)	340.6*** (11.24)
N	474	475

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- Differential Controller Effects

Specification

$$y_{tx}^i = c_0 + \gamma M_t + \beta \text{weekend}_t + \epsilon_t \quad (2)$$

- Lag variable $x \in [1, 30]$.
- Examines how commits on day t affect user behavior in time $t + x$.

Average Activity Logs Per Distinct User

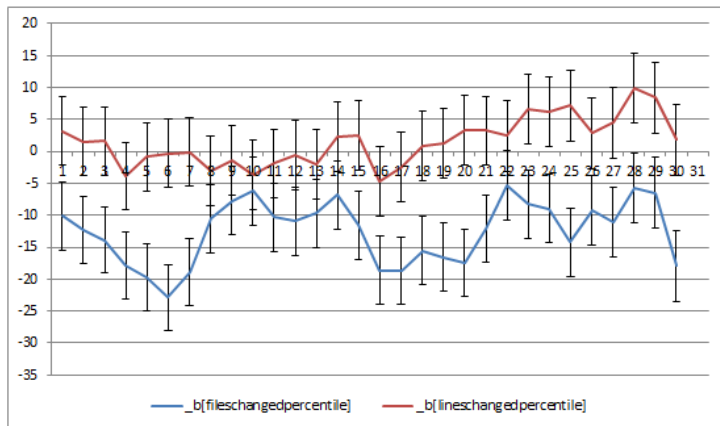


Figure: γ Coefficients with Varying Lags, Regressed on Average Activity Logs per User

Average Event Logs Per Distinct User

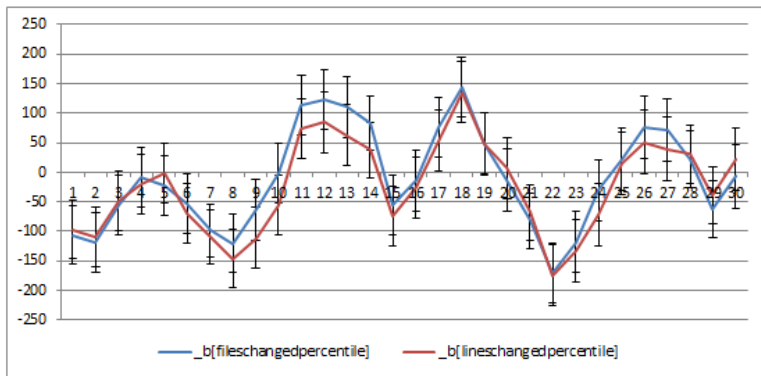


Figure: γ Coefficients with Varying Lags, Regressed on Average Event Logs per User

Insertions and Deletions

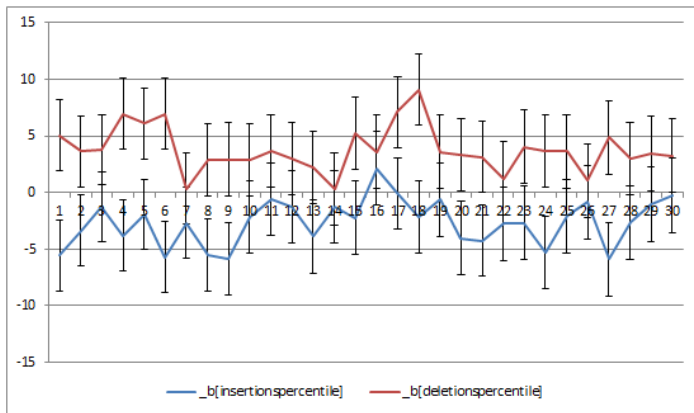


Figure: γ_0 and γ_1 Coefficients with Varying Lags. Regressed on Average Activity Logs per User

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- **Why Micro-Level Data?**
- Search Controller
- Commit Types
- Differential Controller Effects

Problems with Macro-Level Data

- Nothing more than correlations
- Possibly complex, unknown mechanisms for how results come about

The Case for Micro-Level Data

- Panjiva has data on the page and time that any action was performed.
- Code changes (commits) can be thought of as exogenous shocks.
- Almost all changes are unannounced
- Only extremely large changes are announced on blog (less than 1% of Panjiva's total pageviews come from blog).

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- **Search Controller**
- Commit Types
- Differential Controller Effects

Search Controller

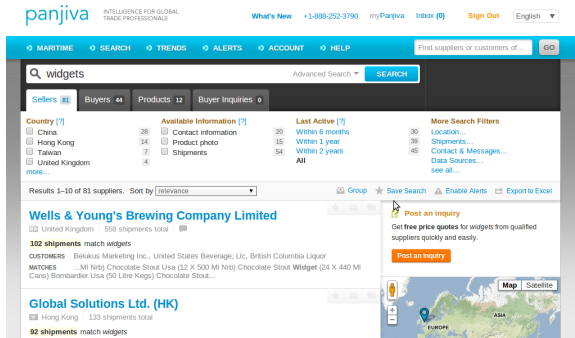


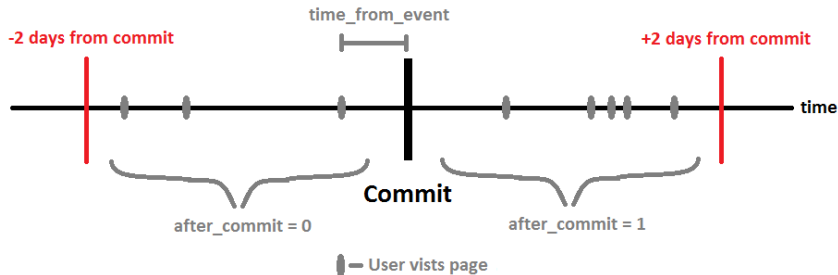
Figure: The Search Page, Panjiva's most trafficked page, provides functionality for finding suppliers, buyers, products, and buyer inquiries.

Specification

$$num_views_day_after_{it} = c_0 + \beta_0 \mu_{it} + \beta_1 hour_dummies_{it} + \epsilon_{it} \quad (3)$$

- i indexes the i th commit and t is the t th view in 4 day window surround the i th commit.
- μ_{it} is either *after_commit* or *time_from_event*.

Visualization of Variables



Search Regression Results

	Dependent Variable: num_views_day_later			
	(1)	(2)	(3)	(4)
after_commit	5.541*** (17.91)		3.378*** (11.20)	
time_from_event		0.0000737*** (43.76)		0.000110*** (66.60)
created_at_hour dummies?	No	No	Yes	Yes
_cons	156.9*** (731.32)	159.9*** (1033.47)	138.3*** (127.71)	139.7*** (130.50)
<i>N</i>	2345617	2345617	2345617	2345617

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

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- **Commit Types**
- Differential Controller Effects

Specification

$$\begin{aligned} num_views_day_later_{it} = & \beta_0 after_commit_{it} + \beta_1 insertionspercentile_{it} + \\ & + \beta_3 insertionspercentile_{it} * after_commit_{it} \\ & + \beta_4 deletionspercentile_{it} * after_commit_{it} \\ & + \beta_5 hour_dummies_{it} + c_0 + \epsilon_{it} \end{aligned}$$

- Interaction term coefficients β_3 and β_4 give incremental percentile coefficients after the commit
- Same specification, but including interaction terms to see additive effect

Differences in Commits

	Dependent Variable: num_views_day_later			
	(1)	(2)	(3)	(4)
after_commit	-10.86*** (-16.90)	-13.79*** (-22.03)	-325.5*** (-154.63)	-305.9*** (-149.28)
insertionspercentile	-50.97*** (-54.12)	-50.76*** (-55.34)	-650.6*** (-218.09)	-609.5*** (-209.88)
deletionspercentile	66.31*** (70.67)	63.35*** (69.34)	251.5*** (84.73)	230.8*** (79.94)
insertionspercentile * after_commit	57.00*** (41.21)	58.05*** (43.09)	393.0*** (89.07)	365.1*** (85.01)
deletionspercentile * after_commit	-24.16*** (-17.65)	-23.64*** (-17.74)	2.306 (0.53)	3.295 (0.77)
Hour Dummies?	No	Yes	No	Yes
Controllers Used	Search	Search	All	All
_cons	150.1*** (336.67)	133.2*** (116.37)	606.5*** (420.98)	491.0*** (130.17)
N	2345617	2345617	3858943	3858943

1 Introduction

- Silicon Valley Mindset
- What We Know
- Questions to Ask
- Panjiva Dataset

2 Macro-Level Results

- Daily Effects of Code Changes
- Lagged Effects of Code Changes

3 Micro-Level Results

- Why Micro-Level Data?
- Search Controller
- Commit Types
- **Differential Controller Effects**

Effects of Each Controller

$$\begin{aligned} num_views_day_later_{it} &= \beta_0 after_commit_{it} + \bar{\beta}_1 \overline{controllers}_{it} \\ &+ \bar{\beta}_2 \overline{controllers}_{it}^T \times \overline{after_commit}_{it} \\ &+ \beta_3 hour_dummies_{it} + \epsilon_{it} \end{aligned}$$

- Looking for the impact of controller k on user activity after a commit
- Want to examine $\Gamma_k = \beta_{2k} + \beta_{1k} + \beta_0 - \beta_{1k} = \beta_{2k} + \beta_0$.
- Standard errors given by:

$$SE_{sum} = \sqrt{SE_{cont_{2k}}^2 + SE_{ac}^2 + 2Cov(cont_{2k}, ac)}$$

Controller Results

	(1)	(2)
	No Hour Controls	With Hour Controls
Communication	-2.673 (-0.50)	5.111 (0.98)
My_Panjiva	-41.718* (-1.98)	-54.530*** (-2.68)
Profile	-8.167 (-1.30)	-6.339 (-1.04)
Project	-410.438*** (-235.55)	-388.496*** (-228.27)
Search	5.541*** (4.37)	1.516 (1.23)
US_Exports	10.318 (0.28)	-28.946 (-0.80)
US_Imports	-12.487 (-0.82)	-8.73 (-0.59)
<i>N</i>	3858943	3858943