

# Data Privacy Paradox and Digital Demand

Wei Xiong

Joint work with Long Chen, Yadong Huang, Shumiao Ouyang

CB&DC Virtual Seminar

December 16, 2021

# Motivation

- Data sharing empowers the booming fintech industry and digital economy
  - Data as a third factor of macroeconomy, e.g., Jones & Tonetti (2020), Farboodi & Veldkamp (2020), Cong, Xie & Zhang (2020)
- Privacy concerns are becoming more and more important
  - General Data Protection Regulation (GDPR), effective on May 25, 2018
  - California Consumer Privacy Act of 2018 (CCPA), effective on January 1, 2020
- **What are consumers' privacy concerns? How do their privacy concerns affect their data sharing?**
- The data privacy paradox, e.g., Acquisti, Brandimarte & Loewenstein (2020)
  - Consumers state concerns about data privacy in surveys, yet they share their personal data freely or for small rewards
- Does the paradox exist in data sharing with digital platforms? If so, why?

# Summary

- We examine this paradox by combining survey and behavioral data of a sample of Alipay users
- We confirm the data privacy paradox
  - Users with stronger privacy concerns authorize data sharing with the same number of mini-programs in Alipay
- What explains the data privacy paradox?
  - Not due to unreliable survey responses or frustration
  - Users with stronger privacy concerns use mini-programs more intensively
  - Heavy users of mini-programs are more likely to cancel data sharing with mini-programs
- Privacy concerns are likely developed as a **by-product** in the process of using digital applications

# Related Literature

- Data privacy paradox, e.g., Gross & Acquisti (2005), Goldfarb & Tucker (2012), Athey et al. (2017)
  - Ignorance about data sharing, e.g., (Pew, 2019)
  - Illusion of control, e.g., Brandimarte, Acquisti & Loewenstein (2013)
  - Present bias, .e.g., Acquisti (2004)
  - Value of data privacy, e.g., Acquisti, John & Lowenstein (2013), Tang (2019)
- Why do consumers care about data privacy?
  - Price discrimination: Taylor (2004), Acquisti & Varian (2005), Ichihashi (2020)
  - Data security: Fainmesser et al. (2019)
  - Protecting personal vulnerability: Liu, Sockin & Xiong (2020)

# The Alipay Platform

- A payment and lifestyle platform with over 900 millions active users in China
- Millions of mini-programs operated by **third parties**, each requiring authorization of data sharing at the initial entry
  - Substantial variations in services and data sharing

Search Box for Mini-Programs

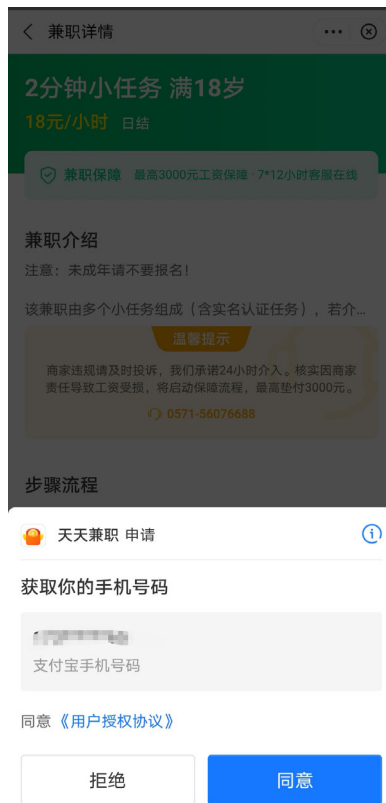
Shortcuts to Frequently Used Mini-Programs

Message Box (Survey Access)



# Examples of Mini-Program Authorization Page

*Part-time  
Job  
Searching*



*Friend  
Making*



*Law  
Consulting*



**Information Required:** *Mobile Number*

*Nickname, Profile, Gender, Area*

*Current Location*

# The Survey

- In July 2020, a survey was sent through the Alipay message box.
  - Over two million Alipay active users randomly chosen
  - 27,597 clicked on the survey link
  - 14,250 completed the survey
  - 10,875 users say they used mini-programs in Alipay
- The survey contains 12 questions
  - Typically completed in 1-3 mins
  - The respondents are well spread across China

Search Box for  
Mini-Programs

Shortcuts to  
Frequently Used  
Mini-Programs

Message Box  
(Survey Access)



# Responses to Some Survey Questions

	Count	Total	Share
<i>A. Are you concerned about privacy issues while using online services?</i>			
Very concerned	13284	14250	93%
Concerned	882	14250	6%
Not concerned	84	14250	1%
<i>B. What do you think about privacy protection in Alipay?</i>			
Very good	6789	14250	48%
Ordinary	5600	14250	39%
Not good	679	14250	5%
No idea	1182	14250	8%
<i>F. Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?</i>			
Very concerned	5005	10875	46%
Concerned	4244	10875	39%
Not concerned	1626	10875	15%
<i>G. What privacy issues are you concerned about when using mini-programs in Alipay? (multiple choices)</i>			
Data leakage and security	9377	10875	86%
Price discrimination by merchants	2314	10875	21%
Seductive advertising and temptation consumption	5333	10875	49%
Others	500	10875	5%

This question is specifically linked to data sharing with mini-programs

- Immune to the concern of Solove (2021)



# Summary Statistics

Panel A: Sample of Survey Respondents

	N	Mean	Std	Min	p25	Median	p75	Max
Part I. General information								
Concerned Dummy	10875	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Very Concerned Dummy	10875	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Privacy Setting Changed	10875	0.49	0.5	0.00	0.00	0.00	1.00	1.00
Digital Experience (month)	10871	74.97	35.07	4.00	48.00	70.00	97.00	190.00
Age (year)	10858	32.82	10.27	10.00	25.00	31.00	39.00	82.00
Part II. Data sharing with mini-programs								
# Authorized Mini-Programs	10875	11.37	7.63	0.00	7.00	10.00	14.00	93.00
# Entered Mini-Programs	10875	15.72	12.06	1.00	10.00	13.00	19.00	275.00
Has Canceled	10857	0.48	0.50	0.00	0.00	0.00	1.00	1.00
# Cancellations	10612	0.26	0.98	0.00	0.00	0.00	0.00	23.00
Cancellation Rate	10612	0.04	0.12	0.00	0.00	0.00	0.00	1.00
Part III. Use of mini-programs								
Monthly Mini-Program Use								
# Active Days	1521645	0.57	2.92	0.00	0.00	0.00	0.00	31.00
# Uses	1521645	0.81	5.01	0.00	0.00	0.00	0.00	75.00
# Launches	1521645	2.29	15.07	0.00	0.00	0.00	0.00	230.00
# Visited Pages	1521645	5.20	33.67	0.00	0.00	0.00	0.00	503.00

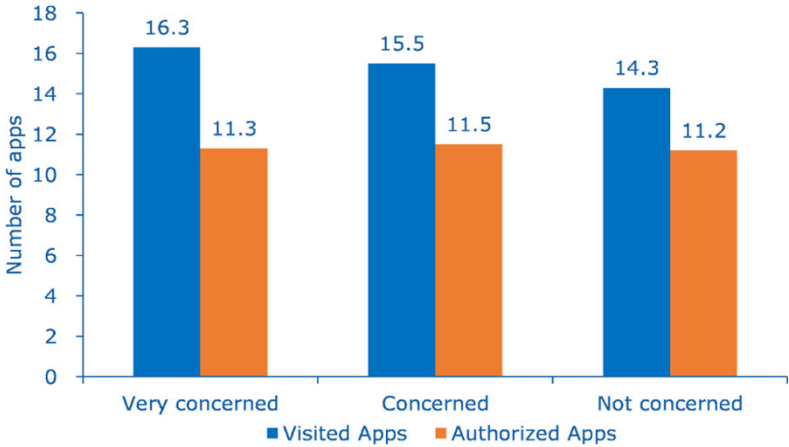
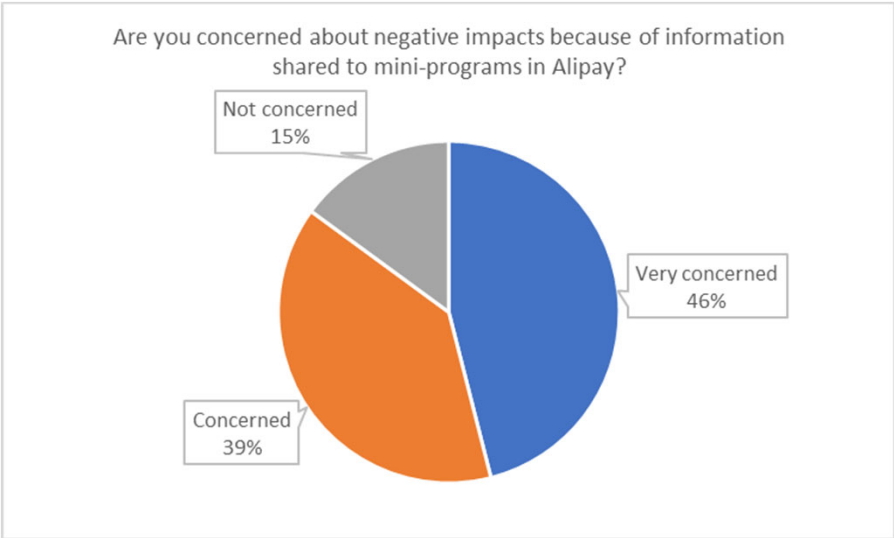
# Simple Framework

- Consider **user  $i$** 's data sharing choice with **mini-program  $j$** :
  - The cost is  $c_{ij} = c_i + c_j + \epsilon_{ij}$
  - The benefit is  $b_{ij} = b_i + b_j + \epsilon_{ij}$
  - The user will authorize if

$$b_{ij} - c_{ij} = b_i - c_i + b_j - c_j + \epsilon_{ij} - \epsilon_{ij} > 0.$$

- **Hypothesis 1**: other things being equal, privacy concerned users are **more reluctant** to authorize data sharing

# The Data Privacy Paradox



# User Level Analysis

$$Y_i = a_1 \text{Concerned}_i + a_2 \text{Very Concerned}_i + a_3 \text{Age}_i + a_4 \text{Digital Experience}_i + \delta_i + \varepsilon_i,$$

	# Authorized Mini-programs		# Visited Mini-programs	
	(1)	(2)	(3)	(4)
Concerned Dummy	0.334 (0.213)	0.207 (0.214)	1.262*** (0.322)	1.243*** (0.320)
Very Concerned Dummy	0.127 (0.209)	-0.007 (0.211)	1.990*** (0.331)	1.965*** (0.336)
Digital Experience		0.012*** (0.002)		-0.002 (0.004)
Age		-0.039*** (0.009)		0.204*** (0.015)
Constant	11.177*** (0.178)		14.310*** (0.274)	
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Observations	10875	10858	10875	10858
Adjusted R2	0.0001	0.021	0.003	0.045

# User-Mini-Program Level Analysis

$$Y_{ij} = a_1 \text{Concerned}_i + a_2 \text{Very Concerned}_i + a_3 \text{Age}_i + a_4 \text{Digital Experience}_i + \delta_i + \gamma_j + \varepsilon_{ij}$$

	Authorized Dummy (0/1)		Visited Dummy (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy ( $\times E-4$ )	0.862 (0.745)	0.386 (0.735)	2.897*** (0.848)	2.552*** (0.836)
Very Concerned Dummy ( $\times E-4$ )	0.028 (0.736)	-0.465 (0.728)	3.755*** (0.846)	3.340*** (0.840)
Digital Experience ( $\times E-6$ )		5.517*** (0.800)		3.806*** (0.960)
Age ( $\times E-5$ )		-1.958*** (0.287)		2.405*** (0.367)
Constant	0.004*** (0.0001)		0.005*** (0.0001)	
Mini-program FE	N	Y	N	Y
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Observations	25414875	25364288	25414875	25364288
Adjusted R2	0.000	0.105	0.000	0.129

# Validating Survey-Based Privacy Concerns

## User Level Analysis

	Has Canceled (0/1)		Privacy Setting Changed (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy	0.060*** (0.014)	0.033*** (0.014)	0.028* (0.015)	0.012 (0.015)
Very Concerned Dummy	0.082*** (0.014)	0.051*** (0.014)	0.060*** (0.014)	0.041*** (0.015)
Digital Experience		0.004*** (0.0001)		0.001*** (0.0001)
Age		-0.003*** (0.0005)		-0.001*** (0.0005)
Constant	0.420*** (0.012)		0.454*** (0.012)	
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Observations	10,857	10,841	10,875	10,858
Adjusted R2	0.003	0.097	0.002	0.011

## Determinants of Data-Sharing Authorization in Survey

	Count	Share	Count	Share	Total
	Agree		Disagree		
<i>Q1: I agree to authorize data sharing with mini-programs because it is safe in Alipay.</i>					
Concerned or very concerned	3918	42%	5331	58%	9249
Not concerned	1308	80%	318	20%	1626
<i>Q2: I agree to authorize data sharing with mini-programs because my information has already been shared in many platforms.</i>					
Concerned or very concerned	1083	12%	8166	88%	9249
Not concerned	493	30%	1133	70%	1626
<i>Q3: I have to share my information in exchange for digital services even though I have concerns about my data privacy.</i>					
Concerned or very concerned	6030	65%	3219	35%	9249
Not concerned	913	56%	713	44%	1626
<i>Q4: I only authorize data sharing with mini-programs when the requested data are not important.</i>					
Concerned or very concerned	1852	20%	7397	80%	9249
Not concerned	485	30%	1141	70%	1626
<i>Q5: I tend to authorize data sharing with mini-programs that are used by my friends.</i>					
Concerned or very concerned	4042	44%	5207	56%	9249
Not concerned	942	58%	684	42%	1626

# Simple Framework

- Consider user  $i$ 's data sharing choice with mini-program  $j$ :
  - The cost is  $c_{ij} = c_i + c_j + \epsilon_{ij}$
  - The benefit is  $b_{ij} = b_i + b_j + \epsilon_{ij}$
  - The user will authorize if

$$b_{ij} - c_{ij} = b_i - c_i + b_j - c_j + \epsilon_{ij} - \epsilon_{ij} > 0.$$

- **Hypothesis 2:** other things being equal, privacy concerned users use mini-programs **less intensively**.

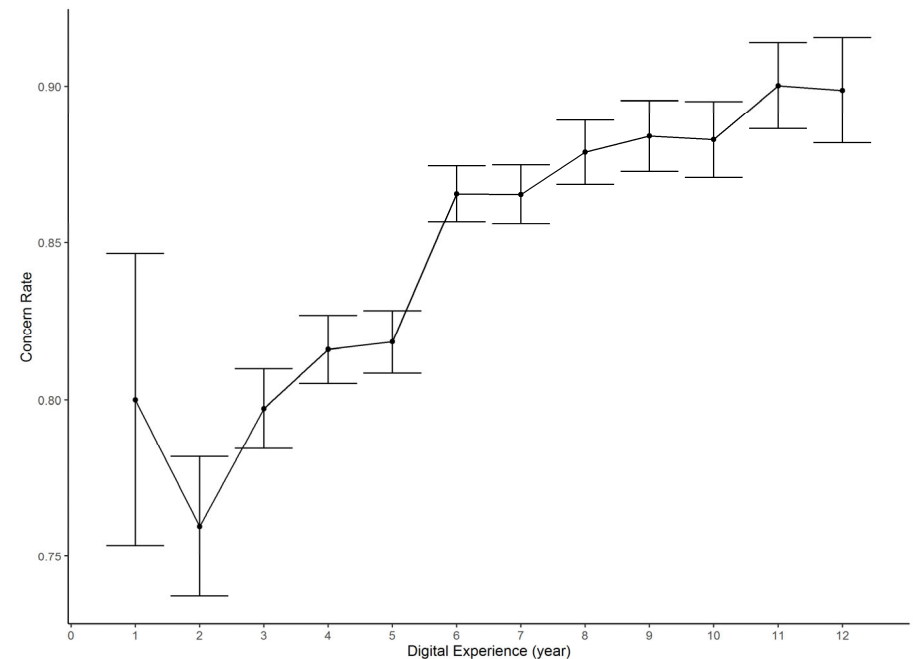


# Demands for Digital Services

	# Active Days		# App Uses		# App Launches		# Visited Pages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concerned Dummy	0.102*** (0.027)	0.088*** (0.020)	0.155*** (0.046)	0.138*** (0.035)	0.434*** (0.131)	0.399*** (0.105)	0.847*** (0.262)	0.772*** (0.219)
Very Concerned Dummy	0.126*** (0.028)	0.102*** (0.021)	0.206*** (0.048)	0.172*** (0.037)	0.568*** (0.135)	0.490*** (0.110)	1.144*** (0.269)	0.996*** (0.230)
Digital Experience		-0.0001 (0.000)		-0.0003 (0.001)		-0.001 (0.001)		-0.001 (0.003)
Age		0.020*** (0.001)		0.033*** (0.002)		0.080*** (0.005)		0.128*** (0.011)
Constant	0.468*** (0.023)		0.651*** (0.039)		1.864*** (0.112)		4.339*** (0.226)	
Mini-program FE	N	Y	N	Y	N	Y	N	Y
Year-Month FE	N	Y	N	Y	N	Y	N	Y
City FE	N	Y	N	Y	N	Y	N	Y
Gender FE	N	Y	N	Y	N	Y	N	Y
Observations	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020
Adjusted R2	0.0002	0.119	0.0002	0.096	0.0001	0.086	0.0001	0.078

# Why Don't Privacy Concerns Deter Digital Demands?

- Users are likely to develop privacy concerns in the process of using digital applications
- **Hypothesis 3:** other things being equal, heavy users of mini-programs are **more likely to cancel** data-sharing with mini-programs



# Activeness and Cancellation

## User Level Analysis

	Cancellation Rate			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.042*** (0.008)		0.080*** (0.016)	
log(1+ # Avg. Monthly Active Sessions)		0.005*** (0.001)		0.012*** (0.003)
Digital Experience ( $\times E-4$ )	-0.112 (0.194)	-0.203 (0.194)	-1.834*** (0.448)	-2.000*** (0.454)
Age ( $\times E-4$ )	-1.250* (0.746)	-0.549 (0.689)	-1.666 (1.896)	-0.682 (1.823)
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	9,860	9,860	3916	3916
Adjusted $R^2$	0.012	0.005	0.027	0.014

# Activeness and Cancellation

## User-Mini-Program Level Analysis

	Canceled Dummy (0/1)			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.038*** (0.007)		0.072*** (0.013)	
log(1+ # Avg. Monthly Active Sessions)		0.003** (0.001)		0.007*** (0.002)
Digital Experience ( $\times E-4$ )	-0.143 (0.173)	-0.221 (0.175)	-1.530*** (0.396)	-1.659*** (0.402)
Age ( $\times E-4$ )	-0.922 (0.636)	-0.097 (0.601)	-0.726 (1.489)	0.451 (1.437)
Mini-program FE	Y	Y	Y	Y
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	64,611	64,611	28,034	28,034
Adjusted $R^2$	0.01	0.009	0.028	0.024

# The Random Sample

- The survey sample is biased toward more active users
- A random sample of 100,000 users from all active Alipay users
  - Robustness checks of the key findings from the survey sample
  - An event study of user reactions to a privacy-related incident

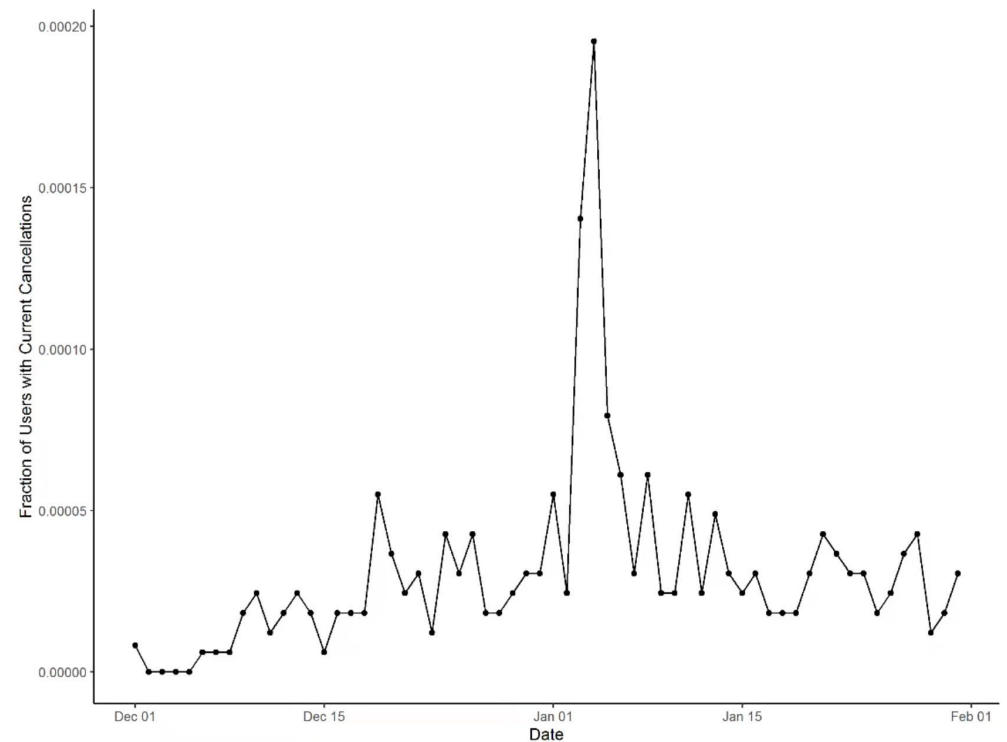
# Summary Statistics

Panel B: Random Sample of Alipay Users

	N	Mean	Std	Min	p25	Median	p75	Max
Part I. General information								
Privacy Setting Changed	98679	0.09	0.28	0.00	0.00	0.00	0.00	1.00
Digital Experience (month)	99600	60.69	36.81	0.00	32.00	55.00	82.00	190.00
Age (year)	97876	36.61	12.89	1.00	27.00	34.00	46.00	120.00
Part II. Data sharing with mini programs								
# Authorized Mini-Programs	100000	2.40	3.52	0.00	0.00	1.00	3.00	136.00
# Entered Mini-Programs	100000	3.02	4.59	0.00	0.00	2.00	4.00	248.00
Has Canceled	99995	0.12	0.32	0.00	0.00	0.00	0.00	1.00
# Cancellations	57214	0.02	0.31	0.00	0.00	0.00	0.00	38.00
Cancellation Rate	57214	0.00	0.05	0.00	0.00	0.00	0.00	1.00
Part III. Use of mini-programs								
Monthly Mini-Program Use								
# Active Days	3036555	0.27	1.59	0.00	0.00	0.00	0.00	27.00
# Uses	3036555	0.34	2.21	0.00	0.00	0.00	0.00	40.00
# Launches	3036555	1.10	6.90	0.00	0.00	0.00	0.00	123.00
# Visited Pages	3036555	3.06	19.96	0.00	0.00	0.00	0.00	342.00

# An Event Study

- On Jan 3, 2018, Alipay launched its Annual User Footprint Report within its mobile wallet app
  - A box consenting to the “Sesame Credit Service Agreement” was set as the default choice on the report's landing page.
  - Users who failed to notice the checked box would have unintentionally agreed to use Alipay’s Sesame credit service.
  - Some internet users quickly noticed this misleading design, and the incident went viral on Chinese social media



# An Event Study

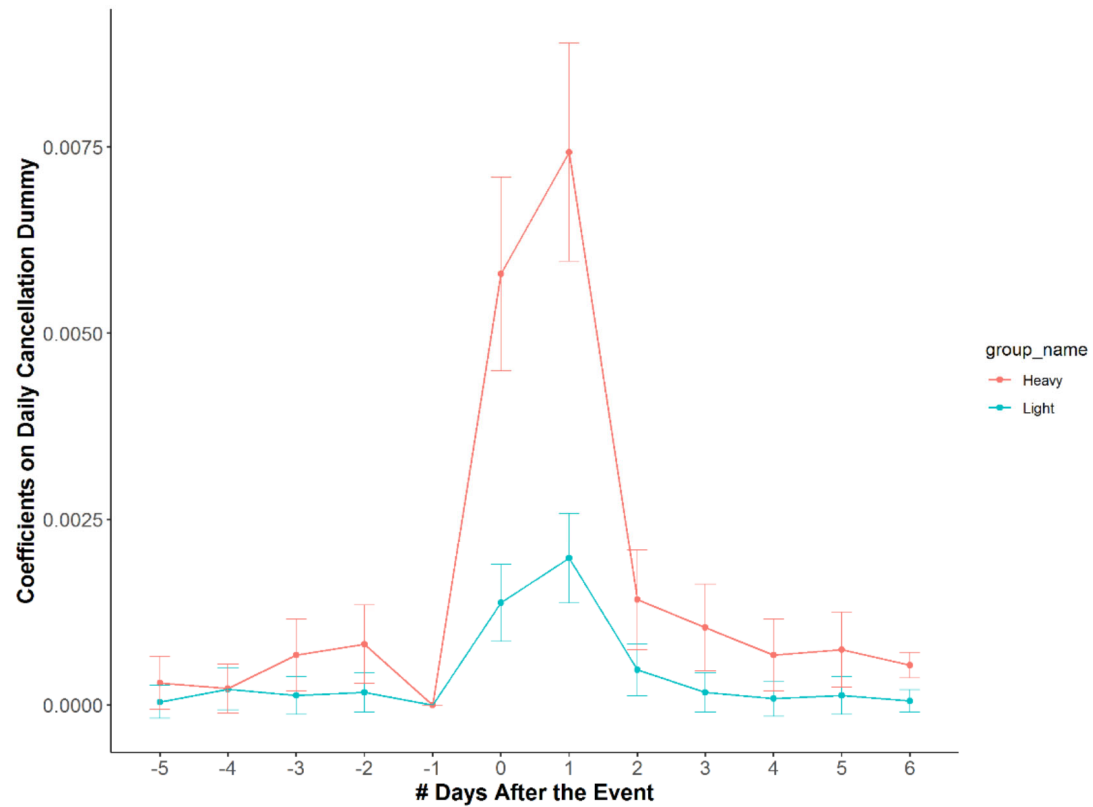
- **Hypothesis 4:** after the incident, heavy users of mini-programs are **more likely to cancel** data sharing with mini-programs

$$\begin{aligned} \text{Daily Cancellation Dummy}_{i,t} = & \alpha_0 + \sum_{\substack{\tau=-5 \\ \tau \neq -1}}^5 \beta_{H,\tau} \cdot \text{Heavy User}_i \cdot \mathbb{1}(t = \tau) \\ & + \beta_{H,6} \cdot \text{Heavy User}_i \cdot \mathbb{1}(t \geq 6) + \sum_{\substack{\tau=-5 \\ \tau \neq -1}}^5 \beta_{L,\tau} \cdot \text{Light User}_i \cdot \mathbb{1}(t = \tau) \\ & + \beta_{L,6} \cdot \text{Light User}_i \cdot \mathbb{1}(t \geq 6) + \delta_i + \varepsilon_{i,t}, \end{aligned} \quad (4)$$



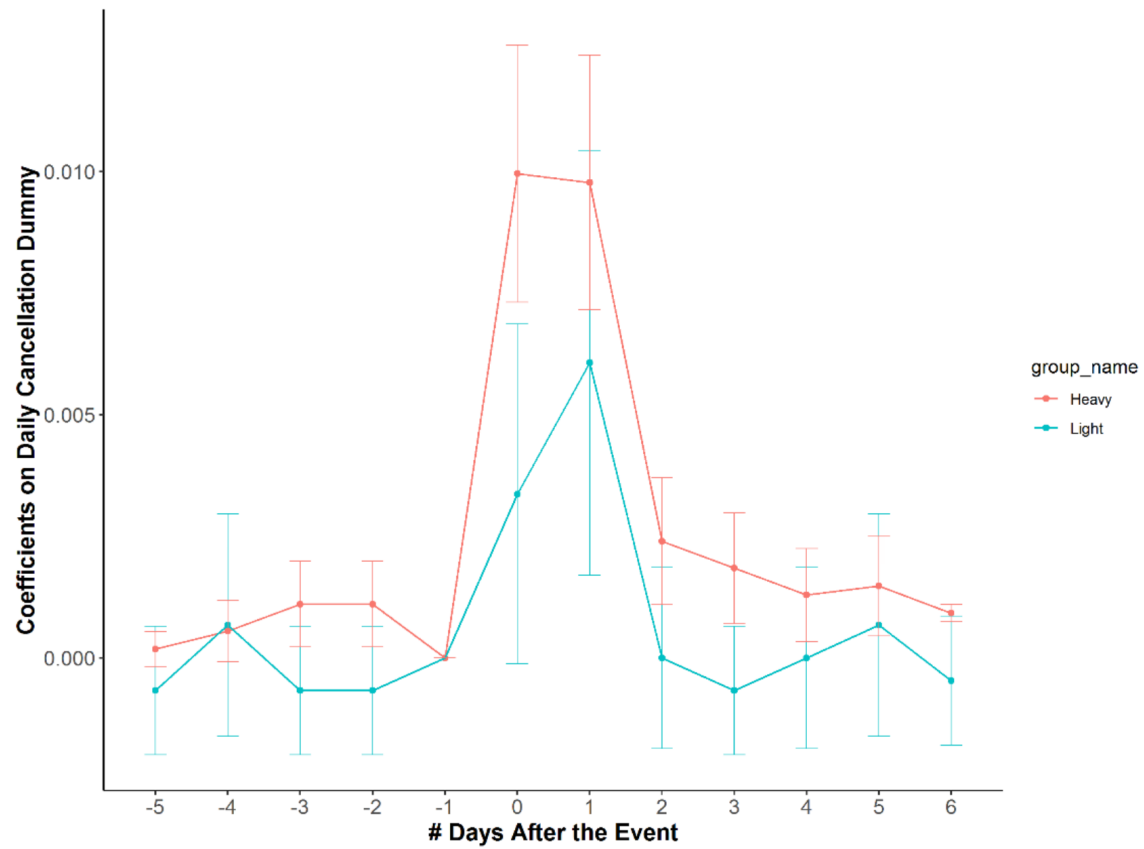
# An Event Study

Unfiltered Users in the Random Sample



# An Event Study

Users with Cancellation before Nov 30, 2017



# Robustness: The Random Sample

Panel A. User Level Analysis of the Data Privacy Paradox

	# Authorized Apps		# Visited Apps	
	(1)	(2)	(3)	(4)
Privacy Setting Changed	2.851*** (0.083)	2.443*** (0.082)	3.599*** (0.117)	3.158*** (0.116)
Controls	N	Y	N	Y
Observations	98,679	96,596	98,679	96,596
Adjusted $R^2$	0.023	0.094	0.022	0.068

# Robustness: The Random Sample

Panel B. User-Mini-Program-Month Level Analysis of Privacy Concerns and Digital Demand

	# Active Days		# Active Sessions		# App Launches		# Visited Pages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Privacy Setting Changed	0.032***	0.043***	0.042***	0.059***	0.102***	0.173***	0.301***	0.521***
	(0.009)	(0.007)	(0.012)	(0.010)	(0.034)	(0.031)	(0.086)	(0.081)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	3,021,210	3,007,635	3,021,210	3,007,635	3,021,210	3,007,635	3,021,210	3,007,635
Adjusted R2	0.00005	0.061	0.00004	0.052	0.00003	0.046	0.00003	0.045

# Robustness: The Random Sample

Panel C. User-Mini-Program Level Analysis of Activeness and Cancellation

	Cancellation Rate			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.006*** (0.001)		0.022*** (0.008)	
log(1+ # Avg. Monthly Active Sessions)		0.002*** (0.0004)		0.008*** (0.003)
Controls	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	57,146	57,146	8,057	8,057
Adjusted R2	0.003	0.002	0.042	0.041

# Summary

- We confirm the data privacy paradox
  - Users with stronger privacy concerns authorize data sharing with the same number of mini-programs in Alipay
- What explains the data privacy paradox?
  - Not due to unreliable survey responses or frustration with protecting privacy
  - Users with stronger privacy concerns also use mini-programs more intensively
  - Heavy users of mini-programs are also more likely to cancel data sharing with mini-programs, both unconditionally and in response to a privacy-related incident
- Privacy concerns are likely developed as **a by-product** in the process of using digital applications
  - Privacy concerns may intensify with the deepening of the digital economy, limiting the increasing returns of data sharing
  - Need to better protect data privacy and thus enable more data sharing

Thank You!