Data Privacy and Fintech

Wei Xiong, Princeton University

Keynote Speech
Midwest Finance Association 2021 Annual Meeting
March 19, 2021

Booming Fintech Industry

- Powered by rapid developments of big data techniques and increasingly accessible consumer data, e.g., Allen, Gu & Jagtiani (2020) and Luohan Academy (2020)
  - Alternative and unconventional data improve credit access and promote financial inclusion

- Internet-based banking and investment services
- Peer-to-peer lending platforms
- Mobile payment and fast payment systems
- Bigtech lending programs

Blockchain Revolution

- Distributed ledger technology, e.g., Townsend (2020)
  - Decentralization of record-keeping
  - Largely popularized by Bitcoin

- Cryptocurrencies and initial coin offerings
  - Over 4000 cryptocurrencies in existence in 2020

- Decentralization of finance

- Central Bank digital currency (CBDC)

Data Privacy

- Data as a third factor of the macroeconomy
  - Jones and Tonetti (2020), Farboodi and Veldkamp (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
  - California Privacy Rights Act of 2020 (CPRA), passed on November 3, 2020

Three related studies

- What are consumers' privacy preferences? How do they make data-sharing choices?
  - Chen, Huang, Ouyang & Xiong (2021) "The Data Privacy Paradox and Digital Demand"

- How do data privacy regulations affect consumers?
  - Liu, Sockin & Xiong (2020) "Data Privacy and Temptation"

- How is the blockchain revolution related to data privacy concerns?
  - Sockin and Xiong (2020) "A Model of Cryptocurrencies"

"The Data Privacy Paradox and Digital Demand"
Chen, Huang, Ouyang & Xiong (2021)

- What are consumers' privacy preferences? How do they make data-sharing choices?

- The data privacy paradox, e.g., Gross and Acquisti (2005), Athey et al. (2017)
  - Consumers state concerns about data privacy in surveys, yet they share their personal data freely or for small rewards

- The data privacy paradox is often interpreted as consumers being not really serious about their data privacy or their survey responses being unreliable

- We examine this paradox by combining survey and behavioral data of a sample of Alipay users

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)

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

Search Box for Mini-Programs

Shortcuts to Frequently Used Mini-Programs

Message Box (Survey Access)

Responses to Some Survey Questions

	Count	Total	Share
Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?			
Very concerned	5005	10875	46%
Concerned	4244	10875	39%
Not concerned	1626	10875	15%
What privacy issues are you concerned about when using mini-programs in Alipay? (multiple choices)			
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%

Summary Statistics

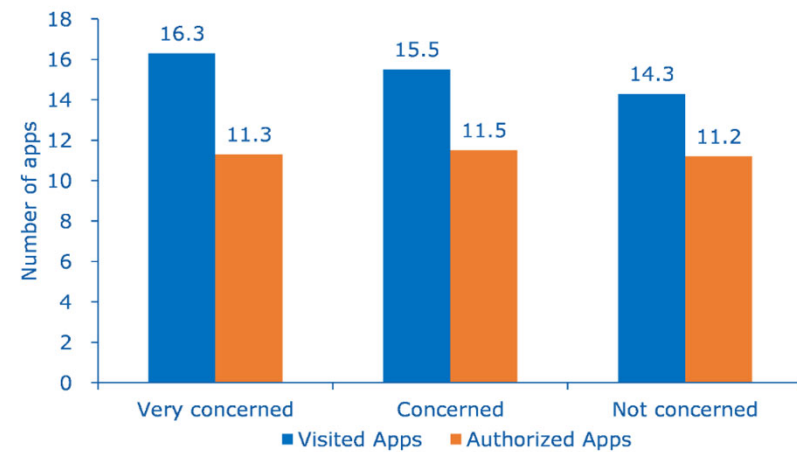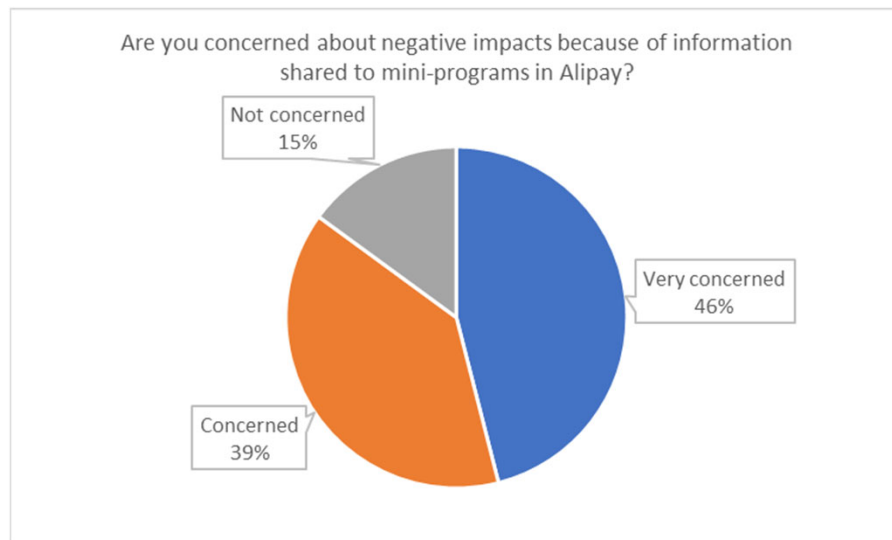|	N	Mean	Std	Min	p25	Median	p75	Max
Part I. General information								
Concerned Dummy	10875	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Very Concerned Dummy	10875	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Privacy Setting Changed	10875	0.49	0.5	0.00	0.00	0.00	1.00	1.00
Digital Experience (month)	10871	74.97	35.07	4.00	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 Cancelled	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. Usage of mini-programs								
Monthly Mini-program Usage								
# 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 $b_{ij} = b_i + b_j + \varepsilon_{ij}$
    - The user will authorize if

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

- Hypothesis: privacy concerned users are more reluctant to authorize data sharing

The Data Privacy Paradox

Are you concerned about negative impacts because of information shared to mini-programs in Alipay?

- Not concerned 15%
- Very concerned 46%
- Concerned 39%

Q3: Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?

The Data Privacy Paradox

User Level Analysis

	# Authorized Mini-programs		# Visited Mini-programs	
	(1)	(2)	(3)	(4)
Concerned Dummy	0.334	0.207	1.262***	1.243***
	(0.213)	(0.214)	(0.322)	(0.320)
Very Concerned Dummy	0.127	-0.007	1.990***	1.965***
	(0.209)	(0.211)	(0.331)	(0.336)
Digital Experience		0.012***		-0.002
		(0.002)		(0.004)
Age		-0.039***		0.204***
		(0.009)		(0.015)
Constant	11.177***		14.310***	
	(0.178)		(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

The Data Privacy Paradox

User-Mini–Program Level Analysis

	Authorized Dummy (0/1)		Visited Dummy (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy (× E-4)	0.862	0.386	2.897***	2.552***
	(0.745)	(0.735)	(0.848)	(0.836)
Very Concerned Dummy (× E-4)	0.028	-0.465	3.755***	3.340***
	(0.736)	(0.728)	(0.846)	(0.840)
Digital Experience (× E-6)		5.517***		3.806***
		(0.800)		(0.960)
Age (× E-5)		-1.958***		2.405***
		(0.287)		(0.367)
Constant	0.004***		0.005***	
	(0.0001)		(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***	0.033***	0.028*	0.012
	(0.014)	(0.014)	(0.015)	(0.015)
Very Concerned Dummy	0.082***	0.051***	0.060***	0.041***
	(0.014)	(0.014)	(0.014)	(0.015)
Digital Experience		0.004***		0.001***
		(0.0001)		(0.0001)
Age		-0.003***		-0.001***
		(0.0005)		(0.0005)
Constant	0.420***		0.454***	
	(0.012)		(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

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 $b_{ij} = b_i + b_j + \varepsilon_{ij}$
  - The user will authorize if

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

- Do privacy concerned users benefit more from using mini-programs?

Demands for Digital Services

	# Active Days		# App Uses		# App Launches		# Visited Pages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concerned Dummy	0.102***	0.088***	0.155***	0.138***	0.434***	0.399***	0.847***	0.772***
	(0.027)	(0.020)	(0.046)	(0.035)	(0.131)	(0.105)	(0.262)	(0.219)
Very Concerned Dummy	0.126***	0.102***	0.206***	0.172***	0.568***	0.490***	1.144***	0.996***
	(0.028)	(0.021)	(0.048)	(0.037)	(0.135)	(0.110)	(0.269)	(0.230)
Digital Experience		-0.0001		-0.0003		-0.001		-0.001
		(0.000)		(0.001)		(0.001)		(0.003)
Age		0.020***		0.033***		0.080***		0.128***
		(0.001)		(0.002)		(0.005)		(0.011)
Constant	0.468***		0.651***		1.864***		4.339***	
	(0.023)		(0.039)		(0.112)		(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?

- Privacy concerns are likely different from risk aversion
  - Risk aversion deters risky investment

- Users are likely to develop privacy concerns in the process of using digital applications

- Are heavy users of digital applications more likely to cancel data-sharing authorizations?
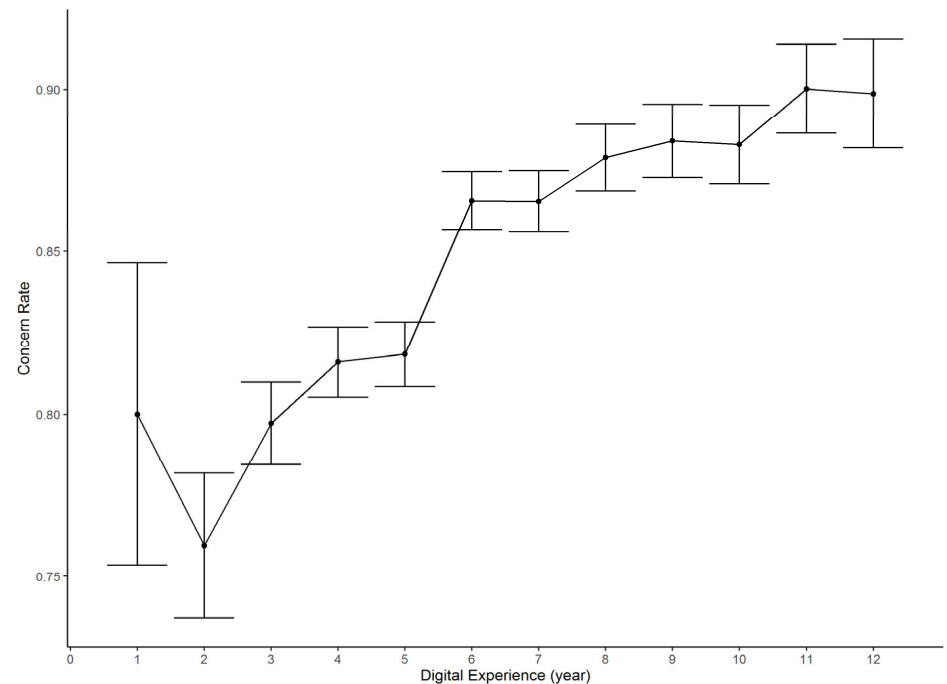
Activeness and Cancellation

User Level Analysis

	Cancellation Rate			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.042***		0.080***	
	(0.008)		(0.016)	
log(1+ # Avg. Monthly Active Sessions)		0.005***		0.012***
		(0.001)		(0.003)
Digital Experience (× E-4)	-0.112	-0.203	-1.834***	-2.000***
	(0.194)	(0.194)	(0.448)	(0.454)
Age (× E-4)	-1.250*	-0.549	-1.666	-0.682
	(0.746)	(0.689)	(1.896)	(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 $R2$	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***		0.072***	
	(0.007)		(0.013)	
log(1+ # Avg. Monthly Active Sessions)		0.003**		0.007***
		(0.001)		(0.002)
Digital Experience (× E-4)	-0.143	-0.221	-1.530***	-1.659***
	(0.173)	(0.175)	(0.396)	(0.402)
Age (× E-4)	-0.922	-0.097	-0.726	0.451
	(0.636)	(0.601)	(1.489)	(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 $R2$	0.01	0.009	0.028	0.024

Summary

- We confirm the data privacy paradox
  - Users with stronger privacy concerns are not more reluctant in authorizing data sharing with mini-programs in Alipay

- What explains the data privacy paradox?
  - Not due to unreliable survey responses
  - Users with stronger privacy concerns also tend to have stronger demands for digital services

- Users are likely to develop privacy concerns 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

"Data Privacy and Temptation"
Liu, Sockin & Xiong (2020)

- A model of privacy preferences through temptation utility
  - Rational consumers always prefer a larger menu, while consumers with weak self-control may suffer from having temptation goods on the menu
  - Data sharing may expose consumers with weak self-control to advertisers and sellers of temptation goods
  - Temptation utility leads to a preference for data privacy, a different approach from price discrimination

- A simple tradeoff of data sharing
  - Improves matching between normal good sellers and consumers
  - Exposes weak-willed consumers to temptation good sellers

- Data sharing comes with positive and negative externalities:
  - Each consumer is affected by data sharing of others, with both improved matching with normal and greater exposure to temptation good

- Welfare ranking of GDPR, CCPA, full sharing, and no sharing

"A Model of Cryptocurrencies"
Sockin & Xiong (2020)

- Decentralization through tokenization to resolve conflicts btw a platform and its users
  - Owners of conventional platforms can't precommit not to abuse control of vast user data
  - As an alternative to equity financing, developers can get paid by issuance of coins and tokens, and delegate controls to a set of precoded algorithms, with no one taking ownership and control

Key Insights

- A tension between network effects and decentralization
  - As users face participation cost, subsidizing users is desirable to maximize network effects
  - A conventional platform has an owner with equity and control to subsidize participation, but the owner cannot pre-commit not to abuse users when profit is low
  - "Trustless" through decentralization makes it possible for a tokenized platform to pre-commit

- Tokenization is appealing when the platform fundamental is relatively weak

Data Privacy & Fintech

- Data sharing and privacy concerns together are likely to reshape the future of the digital economy and the fintech industry

- The data privacy paradox reflects privacy concerns being developed as a by-product in the process of using digital applications
  - Likely to intensify in the coming years

- The social nature of data sharing makes data privacy regulations challenging
  - Privacy preference as derived from temptation utility
  - Positive and negative externalities of data sharing

- Tokenization as an effort to resolve the lack of commitment by conventional platforms to not abuse user privacy