Data Privacy, Data Sharing and Credit Access*

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

July 2025

Abstract

We integrate survey and behavioral data from Alipay to analyze how users' data sharing with third-party mini-programs relates to their privacy concerns, digital engagement, and credit access. Paradoxically, users with stronger privacy concerns do not share less data, exemplifying the data privacy paradox. We resolve this paradox as an omitted variable issue: users with stronger privacy concerns also exhibit a higher demand for digital services offered by mini-programs. Furthermore, our analysis reveals that increased engagement with digital services amplifies privacy concerns. Nevertheless, data sharing not only facilitates access to these services but also enhances credit access.

* Long Chen and Yadong Huang are affiliated with the Luohan Academy, Shumiao Ouyang is affiliated with the University of Oxford, and Wei Xiong is affiliated with Princeton University. We sincerely appreciate the support provided by Yong Li and Dongyu Wei at the Luohan Academy and Xiaopeng Wang and Xiaoqiang Yang at Ant Group. We thank the insightful comments and suggestions from Alessandro Acquisti, Laura Brandimarte, Avi Goldfarb, Zhuang Liu, Cameron Peng, Avanidhar Subrahmanyam, Huan Tang, Catherine Tucker, Liad Wagman, David Yang, and Liyan Yang, as well as seminar participants at the American Finance Association Meetings, Boston College, the CBDC Series, CUHK Shenzhen, EIEF, the IMF, the Kansas City Fed, the NBER Conference on Economics of Privacy, Peking University, Princeton, SAIF, and UCLA. We are particularly grateful to Andreas Fuster and an anonymous referee for their highly constructive feedback in improving this paper. Wei Xiong also thanks the Keynes Fund of the University of Cambridge for financial support. This study has received an exemption from the Institutional Review Board (IRB) of Princeton University.

The rise of FinTech and BigTech lending has made consumer data shared during digital service usage increasingly important to the financial sector. Platforms that initially facilitated transactions, such as mobile payments, now leverage user data for a wide range of financial applications. This repurposing of data—for credit assessments (Berg et al., 2020; Tang, 2020; Gambacorta et al., 2023; Ouyang, 2022), investment recommendations (Rossi and Utkus, 2024), and insurance underwriting (Fang et al., 2020)—highlights its far-reaching implications.[1]

Despite the critical role of data sharing, key questions remain largely unexplored, such as what drives consumers' data-sharing decisions and how these choices impact their access to financial services. This paper addresses these questions using unique data from Alipay. Originally launched in 2004 as an online payment service, Alipay has grown into a comprehensive digital finance ecosystem, offering payments, credit, insurance, and investments. Our dataset integrates users' privacy attitudes, data-sharing choices, digital engagement, and credit access.

The first part of our analysis explores the factors driving consumers' data-sharing behavior. A compelling argument is that data sharing is inversely related to consumers' privacy concerns. The rapid advancements in financial and digital innovations, powered by big data analytics and consumer data sharing, have been accompanied by a significant shift in attitudes toward data privacy. This shift prompted the enactment of major data privacy regulations, including the European Union's General Data Protection Regulation (GDPR) in 2018, California's Consumer Privacy Act (CCPA) in 2020, and China's Personal Information Protection Law (PIPL) in 2021.

To analyze the relationship between consumers' data sharing and privacy concerns, we examined Alipay users' data-sharing authorizations with third-party mini-programs. Introduced in 2017, these lightweight applications operate within the Alipay app, offering services without requiring separate downloads. With over 2 million mini-programs spanning finance, e-commerce, and utilities, third-party providers use Alipay's extensive user base and data infrastructure to deliver personalized services. To access a mini-program, users must consent to share personal data, including initial information and data generated during use. Once authorized, the mini-program continuously collects data, such as transaction details and usage patterns, with the scope of data

[1] Berg, Fuster, and Puri (2022) review the expanding literature on FinTech lending. They highlight how repurposing consumer data for financial applications raises significant privacy concerns while also enhancing access to credit and other financial services. This tension has become more pronounced with the widespread adoption of generative AI in the financial sector. As Mo and Ouyang (2025) emphasize in their review, the transformative use of AI technologies has fundamentally changed how firms leverage consumer data, inevitably intensifying privacy concerns.

sharing varying by transaction type. The diversity of mini-programs—differing in both service value and data sensitivity—provides an ideal setting to examine how users balances privacy concerns against the benefits of digital services.

In July 2020, we surveyed 14,250 Alipay users to assess their concerns about sharing data with mini-programs and linked their responses to administrative data capturing both actual data-sharing behavior and engagement levels. Regarding concerns with sharing personal data with mini-programs: 46% of users reported significant concern, 39% expressed moderate concern, and 15% felt no concern. During our main sample period (July 2019–July 2020), users who were *unconcerned* shared data with an average of 11.2 mini-programs. Surprisingly, *concerned* and *very concerned* users shared with 11.5 and 11.3 mini-programs, respectively.

Contrary to expectations, users with varying degrees of privacy concerns exhibited similar levels of data sharing, even after accounting for user characteristics and mini-program attributes. This finding aligns with the data privacy paradox, a phenomenon previously documented in lab experiments and survey studies by Spiekermann, Grossklags, and Berendt (2001), Gross and Acquisti (2005), Norberg, Horne, and Horne (2007), and Athey, Catalini, and Tucker (2017). The paradox, which highlights the disconnect between privacy concerns and actual behavior, has been a central topic in academic and policy discussions concerning consumer data sharing, e.g., Ben-Shahar (2016), Cooper and Wright (2018), and Fuller (2019). Our findings extend this observation to a real-world context within Alipay's data-driven financial ecosystem.

Notably, administrative data confirms the reliability of self-reported privacy concerns, as users with higher stated concerns were more likely to cancel prior data-sharing authorizations—addressing a critique raised by Solove (2021).

Why do Alipay users with privacy concerns seemingly overlook these apprehensions when permitting data sharing? Our study offers a novel perspective that departs from existing arguments, which suggest that consumers have surrendered their data privacy due to a lack of awareness or behavioral biases, as reviewed by Acquisti, Brandimarte, and Loewenstein (2020). Instead, we attribute the observed privacy paradox to an omitted variable problem: users with greater privacy concerns also tend to have stronger demands for digital services.

Notably, after controlling for users' engagement with mini-programs—which serves as a proxy for their demand for digital services—the relationship between data-sharing authorizations and

privacy concerns turns negative. This result aligns with common expectations, effectively resolving the data privacy paradox.

This surprising role of the omitted variable underscores the importance of understanding how digital engagement may intensify privacy concerns, which we explore in the second part of our analysis. Greater reliance on digital applications allows service providers to accumulate vast amounts of user data, heightening the risks of data breaches (Fainmesser, Galeotti, and Momot, 2022), price discrimination (Taylor, 2004; Acquisti and Varian, 2005), and the exploitation of personal vulnerabilities (Liu et al., 2020), all of which amplify users' privacy concerns.

To establish causal evidence, we use an instrumental variable approach to isolate exogenous shifts in digital demand. Specifically, we instrument for digital engagement using the number of Alipay-bundled shared bikes available in a user's city. This builds on Ouyang (2022), who highlights that the placement of shared bicycles across cities serves as an exogenous factor unrelated to individual privacy concerns. The rationale is that greater availability of shared bicycles encourages more residents to use Alipay for bike access, which in turn deepens their engagement with the broader Alipay ecosystem, including its mini-programs. By leveraging this exogenous variation, we identify a significant causal link between digital engagement and heightened privacy concerns.

We also analyze how privacy concerns evolved following a notable incident on January 3, 2018, which significantly raised awareness of data privacy among Alipay users. Notably, this event made users with deeper engagement in mini-programs more susceptible to heightened privacy concerns, reinforcing the conclusion that increased digital engagement amplifies privacy concerns.

The third part of our analysis presents causal evidence on how data sharing enhances Alipay users' credit access. Using panel data from a representative sample of Alipay users, combined with credit access data from Ouyang (2022), we find that authorizing data sharing with more mini-programs increases credit access, while frequent cancellations of sharing result in reduced credit opportunities. These effects are validated through a 2SLS strategy leveraging time-varying bike-sharing coverage as an instrument for data sharing, isolating exogenous variation in digital engagement. The findings establish a direct causal link between user-authorized data sharing and financial inclusion, particularly in facilitating entry into the credit market.

In summary, our analysis reveals three key insights about data sharing on a major financial platform. First, users share personal data with third-party service providers despite privacy

concerns, driven by their demand for digital services. Second, privacy concerns intensify as digital engagement increases. Finally, data sharing plays a significant role in enhancing users' access to credit.

Our findings contribute to the rapidly expanding literature on the economic implications of consumer data sharing. Theoretical frameworks developed by Jones and Tonetti (2020), Farboodi and Veldkamp (2021), and Cong, Xie, and Zhang (2021) highlight the pivotal role of consumer data sharing in driving financial innovation and sustaining the growing data economy, positioning it as a cornerstone of the broader macroeconomy. For an in-depth review of data sharing within the evolving data economy, see Chen et al. (2021).

The critical role of consumer data in financial services underscores a major trend in fintech: the increasing reliance on consumer data is met with growing privacy concerns and evolving regulations, a dynamic that is set to reshape the industry. Using a survey sample, Armantier et al. (2021) show that U.S. consumers trust banks more than digital platforms when sharing personal data. Meanwhile, Armantier et al. (2024) document significant gender and age differences in data sharing, with female and older individuals being more reluctant to share personal data. Bian et al. (2023, 2021) examine the impact of privacy regulations and disclosures on financial fraud and firm valuation in the mobile app market. Ramadorai, Uettwiller, and Walther (2024) analyze U.S. firms' privacy policies, finding significant variation and a positive correlation between data extraction practices and policy complexity. Agarwal et al. (2024) show that while data breaches negatively affect digital payment platforms, these effects are often short-lived as the convenience of digital services outweighs security concerns. Babina et al. (2025) explore how data regulations shape the emerging field of open banking, while Buchak et al. (2018) highlight technological innovations as a key driver of consumer demand for financial services.

The paper is structured as follows: Section I provides the institutional background of Alipay and its data sharing framework. Section II describes the user survey and summary statistics. Section III analyzes the data privacy paradox, while Section IV examines the relationship between privacy concerns and digital engagement. Section V explores the relationship between data sharing and credit access, and Section VI concludes. Additional analyses are included in the Appendix and a separate Online Appendix.

I. Institutional Background

Alipay, operated by Ant Group, has grown from a payment service on Taobao into the world's largest mobile payment platform. It now serves over 900 million users in China, covering more than 70% of the population. Beyond payments, Alipay offers a wide range of financial services, including small loans, credit cards, insurance, and wealth management. This section provides an overview of Alipay's data-sharing framework.

Mini-Programs

Alipay has expanded its ecosystem through mini-programs, subapplications within the main app that allow third-party providers to offer a wide range of financial and non-financial services—such as wealth management, bike-sharing, and food delivery—without requiring separate app downloads. By June 2020, Alipay hosted over two million mini-programs, with user engagement rising from 21% in Q4 2015 to 49% in Q2 2019 (Chen et al., 2021).

Before Alipay integrated bike-sharing services, consumers relied on standalone apps like Ofo and Mobike, which required separate accounts and payments. Alipay's integration streamlined this process, allowing users to unlock and pay for bikes directly through mini-programs with seamless payments and instant identity verification. Between 2017 and 2019, the bike-sharing industry consolidated as leading apps partnered with platforms like Alipay and WeChat Pay. This integration also enabled researchers to use changes in bike-sharing availability as natural experiments to study digital payment adoption and demand for digital services (Ouyang 2022).

Using mini-programs requires users to exchange personal data for access to digital services. Alipay enforces strict platform-level controls over data practices, requiring explicit user consent for any data shared with a mini-program. Upon first visit, a mini-program requests user authorization to share minimum information necessary for its functionality, following the principle of data minimization.[2] The scope of requested data varies, from basic details like an Alipay nickname to sensitive information such as a national ID number.

[2] For example, Hellobike, a popular bike-sharing mini-program, can be accessed via its standalone app or within Alipay. On first visit, users are prompted to authorize three types of data sharing: (1) basic information like nickname, profile picture, gender, and location; (2) credit score to assess trustworthiness and determine deposit requirements; (3) identification and transaction data, including real name, phone number, national ID, and transaction history. Once authorized, users gain access to Hellobike's services. Figure B1 in the Online Appendix provides three additional examples: a job search mini-program requests mobile numbers, a social connections mini-program requires basic profile and location data, and a legal consulting mini-program requests location access.

Platform policies restrict data collection to operational needs within specific service categories, and sensitive data requests must be standardized and justified. For example, nonfinancial mini-programs (e.g., retail or lifestyle services) may request less sensitive information, like nickname or profile picture, for social features. User consent is mandatory, and service providers must justify additional data requests. Alipay's data policy prohibits mini-programs from using user data for product customization or price discrimination without explicit user consent. Personalization or pricing based on user data must be transparently disclosed and individually authorized at the time of the data request.

Once authorized, mini-programs can continuously collect newly generated data, including transaction details and usage patterns. Users must accept or reject the request, with full access granted only upon authorization.[3] Users can revoke authorization at any time, but doing so typically results in losing access to the service, as continued use depends on ongoing data sharing.

Financial Services

Alipay exemplifies how digital platforms leverage user data—initially collected from digital payments—to expand into diverse financial services such as wealth management, insurance, and credit. This model illustrates how digital footprints across services enable BigTech firms to assess creditworthiness and provide financial services (Berg et al., 2020). Unlike traditional financial institutions that rely primarily on structured credit and transaction records, Alipay centralizes a wide range of user data, including payment, behavioral, and in-app activity data. This distinction makes our finding—that data sharing improves credit access—particularly relevant to platform-based models. As banks adopt open banking and digital services, the implications for user choice, consent, and outcomes grow increasingly significant.

Alipay collaborates with financial partners (e.g., banks, funds, insurance companies) to offer services like virtual credit cards (Huabei), consumer loans (Jiebei), and insurance. To enable these, Alipay shares user data for risk assessment and generates credit scores to determine eligibility for features such as Buy Now, Pay Later and no-deposit rentals. High-scoring users benefit from perks

[3] This setting offers a simpler trade-off compared to data-sharing decisions on many public websites. Under GDPR, websites allow users opt in or out of data collection. If users consent, their data enables personalized services, while opting out still allows access without personalization. Here, the trade-off is between personalized and non-personalized services. In contrast, Alipay users must authorize data sharing to access any mini-program services. Without authorization, they cannot use the mini-program, making the decision a binary choice rather than a spectrum of personalization.

like deposit-free bike rentals, while lower-scoring users must pay deposits. Mini-programs receive ongoing credit score updates unless users revoke authorization.

Risk assessment uses multiple data sources, including Know Your Customer (KYC) identity information, transaction data, and behavioral data from mini-programs. Alipay's extensive behavioral data collection, enabled by its diverse mini-program ecosystem, allows for precise user profiling and enhanced risk assessment.

Recent studies (Huang et al., 2020; Gambacorta et al, 2023) use individual-level data from Alipay to show that digital footprints significantly improve credit risk assessment. Alipay's ability to integrate and analyze diverse data sources underlines the value of its platform-based model in financial services.

Alipay Default Settings

In Alipay's main app, users manage data-sharing settings, such as displaying their real name to friends, making posts public, allowing connections without permission, and being searchable by phone number. These options let users customize privacy preferences, though default settings prioritize visibility and connectivity. Some users adjust these defaults, reflecting privacy concerns about sharing personal information. In our analysis, changes to default settings serve as a behavioral indicator of privacy concerns, complementing our survey-based measure.

II. Survey and Administrative Data

This section describes the survey of Alipay users on privacy concerns and presents summary statistics on data sharing and other administrative data of the survey respondents.

A. The Survey

In July 2020, we collaborated with Alipay to conduct a survey on users' preferences regarding data sharing with third-party mini-programs. The survey comprised 12 questions and was distributed through the message box at the center of Alipay's front page, a highly visible channel, to a random sample of active users.

A total of 27,597 users opened the survey link, and 14,250 completed it. Midway through the survey, respondents were asked, *"Have you ever used mini-programs in Alipay?"* Only those who answered *"yes"* proceeded to the remaining questions on privacy concerns related to data sharing

with mini-programs. Among the respondents, 10,875 reported having used mini-programs, representing 76% of the total.[4] These 10,875 users form the main sample for our analysis.

Given that more-active users are naturally more likely to notice and engage with the message box in the Alipay app, the survey respondents are more representative of active users rather than the entire user base. For robustness and comparison, the Appendix analyzes a separate representative sample of 100,000 Alipay users, randomly drawn from the full population of Alipay users.

The survey was conducted in Chinese, and an English translation of the questions is provided in the Online Appendix. When asked a general question—*"Are you concerned about privacy issues while using digital services?"*—93% of respondents reported being *very concerned*, 6% were *concerned*, and only 1% were *not concerned*. However, when asked specifically about data sharing with mini-programs in Alipay—*"Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?"*—46% were *very concerned*, 39% were *concerned*, and 15% were *not concerned*.

The lower level of concern in the second response, compared to the more general privacy question, aligns with Solove (2021), who emphasizes the need to match privacy concerns with specific data-sharing contexts when analyzing the data privacy paradox. As this second question directly relates to our study, we use respondents' answers as a key measure of privacy concerns in our analysis. Specifically, we construct two dummy variables based on their responses: The *Concerned Dummy* is assigned a value of 1 if a respondent selected *"concerned"* and 0 otherwise. Similarly, the *Very Concerned Dummy* is assigned a value of 1 if the respondent selected *"very concerned"* and 0 otherwise."

Figure 1 categorizes users based on the length of time they have been using Alipay, ranging from one to twelve years, and measures privacy concerns as the percentage of users who report being "*concerned*" or "*very concerned*" about data sharing. It shows that as users gain more experience with digital services, their privacy concerns intensify.

[4] Figures B2–B5 in the Online Appendix provide some characteristics of the survey respondents. It took most respondents more than sixty seconds to complete the survey, indicating that they answered the questions in a serious way (Figure B2). The geographical distribution of the respondents across the provinces in China lines up well with the distribution of the population (see Figure B4), except that the share of respondents from the most populated Guangdong province is about 17%, substantially higher than its population share of about 8.2%.

We also asked respondents a specific question: "*What privacy issues are you concerned about when using mini-programs in Alipay?*" This question allowed multiple selections from four options: (1) data leakage and security, (2) price discrimination by merchants, (3) seductive advertising and temptation consumption, and (4) others. The first option reflects concerns about inadequate protections against hacking and data breaches, as modeled by Fainmesser, Galeotti, and Momot (2022). The second represents the risk that extensive data sharing enables merchants to infer consumers' reservation prices and engage in price discrimination—a concern widely analyzed in the digital economy literature (Acquisti, Taylor, and Wagman, 2016; Bergemann and Morris, 2019; Goldfarb and Tucker, 2019). The third option highlights a growing concern in the digital economy: data sharing may expose consumers' behavioral vulnerabilities, such as a lack of self-control, to advertisers and sellers, as emphasized by Liu, Sockin, and Xiong (2020).

Among respondents, 86% selected data leakage and security, 49% selected seductive advertising and temptation consumption, and 21% selected price discrimination by merchants. Since only 5% selected "others", the first three concerns appear to capture the primary privacy concerns of Alipay users. In our analysis, we also use these responses to measure specific types of privacy concerns.

Regarding privacy controls, we also asked: "*Do you know how to change privacy settings in Alipay?*" and "*Have you ever changed your privacy settings in Alipay?*" 60% of respondents reported knowing how to change their settings, while 39% had actively modified them.[5]

B. Administrative Data

A key strength of our study is access to respondents' extensive administrative data within Alipay, enabling us to analyze the relationship between privacy concerns, actual data-sharing choices, and usage of authorized mini-programs. Table 2 presents summary statistics of key variables. Panel A includes three categories of user information: general profile, data-sharing behavior with mini-programs, and monthly mini-program usage.

[5] Table B2 in the Online Appendix also provides a summary of responses to five additional survey questions: *(1) Will you avoid visiting mini-programs in Alipay because of privacy concerns? (2) How many times will you agree if making authorization decisions for ten mini-programs? (3) How often do you regret authorizing information to mini-programs in Alipay? (4) Do you know how to opt out from mini-programs in Alipay? (5) Have you ever opted out from mini-programs in Alipay?* Additionally, as reported in Section III.A, we compare respondents' answers to three questions with their actual behavior to validate their survey responses.

For general information (user profiles), we have data on each user's gender, age, and city, along with digital experience, measured by the number of months since the user first registered on Alipay. The average user age is 32.82 years, and the average digital experience is 74.97 months.

The data on mini-program data sharing consists of five user-level variables. The first two measure users' data-sharing behavior from July 2019 to December 2021, a period that includes the survey conducted in July 2020. The first variable records how many times users authorize data-sharing requests. The second counts the number of initial visits to mini-programs, which trigger these requests.

The remaining two variables capture cancellations of previously authorized data sharing. As noted earlier, Alipay users can revoke data-sharing permissions at any time. To measure this, we define the $Has\ Canceled_i$ dummy, which takes a value of 1 if a user had canceled data sharing with at least one mini-program between January 2013 and July 2020 (a seven-year period before the survey) and 0 otherwise. The variable $\#\ Cancellations_i$ counts the number of active mini-programs from which a user had revoked data-sharing authorization during this period, where a mini-program is considered active if the user had engaged with it at least once.

In our survey sample, respondents initially visited an average of 46.57 mini-programs between July 2019 and December 2021, with a standard deviation of 55.45 and a maximum of 1,609. The number of data-sharing authorizations averaged 34.22, with a standard deviation of 22.78 and a maximum of 422. These figures indicate that, on average, respondents rejected 26.5% of data-sharing requests. This nontrivial rejection rate suggests that users had not passively accepted all requests but instead exercised discretion in their data-sharing decisions.

Between January 2013 and July 2020, 48% of respondents canceled at least one data-sharing authorization. Despite nearly half of users actively revoking access at some point, the average number of cancellations was 2.66, and the average cancellation rate was 0.05. This low cancellation rate suggests that Alipay users revoked data-sharing authorizations relatively infrequently.

The data on mini-program usage tracks monthly activity at the user × mini-program × month level from July 2019 to July 2020. It includes three key variables: the number of sessions, launches, and page visits. These measures capture different aspects of engagement. A user may interact with a mini-program across multiple sessions in a month, launch it several times per session, and visit

multiple pages within each launch. On average, each month, a respondent in our survey sample was engaged with a mini-program in 0.81 sessions, initiated 2.29 launches, and viewed 5.20 pages.

Panel B of Table 2 compares three user groups—*unconcerned, concerned*, and *very concerned*, across four dimensions: digital experience, age, gender, and education. Notably, both *concerned* and *very concerned* users tend to have longer digital experience, are more likely to be female, and have a higher likelihood of holding a college degree or higher. However, these groups do not exhibit any significant differences in age.[6]

III. The Data Privacy Paradox

This section examines the relationship between respondents' data-sharing choices and their privacy concerns by combining survey responses with administrative data. Figure 2 compares the number of data-sharing authorizations by Alipay users, categorized by their responses to the survey question: "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" The analysis separates data-sharing activity into the pre-survey period (July 2019 – July 2020) and the post-survey period (August 2020 – December 2021).

During the pre-survey period, "unconcerned" users authorized data sharing with an average of 11.2 mini-programs, compared to 11.5 by "concerned" users and 11.3 by "very concerned" users. Despite expressing stronger privacy concerns, "concerned" and "very concerned" users authorized data sharing at levels similar to "unconcerned" users. In the post-survey period, data-sharing activity increased across all user groups. On average, "unconcerned" users shared data with 22.5 mini-programs, compared to 24.6 by "concerned" users and 23.8 by "very concerned" users. Notably, "concerned" and "very concerned" users showed a greater increase in data-sharing activity than their "unconcerned" counterparts.

Taken together, Figure 2 highlights a paradoxical pattern: there is no apparent relationship between the number of data-sharing authorizations and stated privacy concerns during the pre-survey period, while a positive relationship emerges in the post-survey period. We will

[6] A recent study by Armentier et al. (2024) analyzes the privacy concerns of a sample of U.S. consumers, reporting significant differences by gender and age—female consumers and older individuals are significantly more likely to be concerned about data sharing. Similarly, our sample also reveals a significant gender gap in privacy concerns. However, unlike their findings, we do not observe a significant age gap. Instead, our data highlight a significant difference based on digital experience, with users who have longer digital experience being more concerned about data sharing. It is possible that in their sample, age partially captures differences in digital experience.

systematically examine this relationship in our regression analysis. This initial pattern aligns with the data privacy paradox documented in the literature, such as Spiekermann, Grossklags, and Berendt (2001), Gross and Acquisti (2005), Norberg, Horne, and Horne (2007), and Athey, Catalini, and Tucker (2017). While these prior studies are largely based on lab experiments and surveys, our findings extend the analysis to a highly relevant real-world context on a major financial platform.

As this paradox has been a focal point in academic and policy discussions on consumer data sharing, it provides a natural entry point for studying Alipay users' data-sharing behavior. This section analyzes the paradox from three key perspectives. First, we show that the paradox cannot be attributed to spurious privacy concerns expressed in the survey. Second, we use panel regressions at both the user-month and user–mini-program levels to systematically analyze the relationship between privacy concerns and data-sharing authorizations. Finally, we demonstrate that the paradox stems from failing to account for users' underlying digital demands. After controlling for digital demands, we find a robust negative relationship between data sharing and privacy concerns, effectively resolving the paradox.

A. Validating Survey-Based Privacy Concerns

A common critique of the data privacy paradox is that it may reflect inconsistencies in how respondents express privacy preferences rather than a genuine paradox. Skepticism toward survey-based findings is well documented (Bertrand and Mullainathan, 2001), and Solove (2021) similarly questioned whether self-reported privacy concerns in paradox studies align with actual concerns.

We address this concern in two ways. First, we validate survey responses by comparing survey-reported authorization rates, privacy setting changes, and data-sharing cancellations (Table B2 in the Online Appendix) with actual behaviors from administrative data. As shown in Table B3, correlations across these dimensions are positive and highly significant, confirming strong alignment between self-reported and actual behaviors.[7]

Second, we examine whether survey-based privacy concerns correlate with concrete actions to protect data, specifically canceling previously authorized data sharing with mini-programs and

[7] The correlations between survey responses and observed behaviors across these three dimensions are 0.165, 0.183, and 0.192, respectively—all positive and highly significant.

modifying Alipay's default privacy settings. Conceptually, users with stronger privacy concerns should be more likely to take these actions.

Our analysis includes both user-level and user–mini-program-level regressions. For the user-level analysis, we use the following cross-sectional regression model:

$$Y_i = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \epsilon_i\ , \quad (1)$$

where the dependent variable Y_i measures user i's privacy protection behavior, including actions like canceling data sharing with mini-programs or modifying Alipay's default privacy settings. Specifically, Y_i is an indicator for whether a user revoked any data-sharing authorization from January 2013 to July 2020 or changed Alipay's privacy settings between May 2017 and April 2020.[8]

The dummy variable $Concerned_i$ equals 1 for "*concerned"* users, while $Very\ Concerned_i$ equals 1 for "*very concerned*" users. As shown in Table 1, only 60% of survey respondents knew how to modify Alipay's default settings. To account for differences in digital literacy, we include Age_i and $Digital\ Experience_i$ as controls, along with potential fixed effects for city and user gender to capture heterogeneity across these dimensions. Standard errors are clustered at the individual level. Because the analysis uses a cross-sectional regression with one observation per individual, clustering at the individual level is equivalent to applying heteroskedasticity-robust standard errors.

Table 3, Panel A presents the regression results. Columns (1) and (2) use the *Has Canceled*$_i$ indicator as the dependent variable. Controlling for other factors, "*concerned*" and "*very concerned*" users were significantly more likely to revoke permissions for at least one mini-program compared to "*unconcerned*" users. with the "*very concerned*" group showing an even higher likelihood of cancellation.

Columns (3) and (4) focus on the *Privacy Setting Changed*$_i$ indicator. Without controls, users with stronger privacy concerns were more likely to modify Alipay's default settings. However,

[8] Alipay started to record these variables at different points in time, leading to their different periods of measurement. Unfortunately, we cannot observe the exact timing of when the privacy setting was changed in our data.

after adding controls in column (4), only the "*very concerned*" group remains significantly more likely to adjust their settings, while the effect for the "*concerned*" group weakens.

Additionally, both digital experience and age strongly correlate with privacy-protective actions, suggesting that younger and more digitally experienced users are better equipped to manage their privacy settings.

Panel B of Table 3 extends the analysis to the user–mini-program level, focusing on data-sharing cancellations. This granular approach controls for mini-program-specific effects, offering deeper insights into how privacy concerns influence users' likelihood of revoking data-sharing agreements for the same mini-programs.

We expand the regression analysis to all user–mini-program pairs in the sample:

$$Y_{ij} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \gamma_j + \epsilon_{ij}\ . \quad (2)$$

For every user i and mini-program j, the dependent variable Y_{ij} equals 1 if the user canceled data sharing with mini-program j, and 0 otherwise. Similar to Equation (1), Age_i , $Digital\ Experience_i$, and δ_i represent user-level controls. This specification also includes mini-program fixed effects γ_j to account for differences across mini-programs.

Additionally, we incorporate specific privacy concerns reported in the survey—data leakage and security (*Concern Type 1*), price discrimination (*Concern Type 2*), and seductive advertising (*Concern Type 3*). This categorization allows us to assess which concerns are most predictive of cancellation behavior. The analysis covers all active data-sharing consents from January 2013 to July 2020, with a sample size of 480,542 observations.[9]

The results show that "*very concerned*" users are significantly more likely to cancel data-sharing authorizations. Columns (2) and (3) reveal that specific privacy concerns also positively correlate with cancellation activity, with data leakage showing the strongest effect. Concerns about price discrimination and seductive advertising are positive but somewhat weaker. Importantly, including these variables does not materially change the coefficient on "*very concerned,*" suggesting that general apprehension about negative impacts goes beyond any single privacy concern.

[9] Note that this sample includes only user–mini-program pairs that users have previously authorized, resulting in a smaller sample size compared to Panels B of Table 4, which consider all possible user–mini-program pairs.

Overall, Table 3 confirms that survey-based privacy concerns align with actual privacy-protective actions, addressing skepticism about the reliability of self-reported privacy attitudes. These findings suggest that Solove's (2021) concerns about spurious privacy preferences are not applicable here, reinforcing that the data privacy paradox is not a result of measurement error in privacy concerns.

B. Privacy Concerns and Data Sharing

We now examine the relationship between privacy concerns and data sharing using a standard cost-benefit analysis framework. When deciding whether to share personal data with a specific mini-program, Alipay users weigh the benefits they receive against potential privacy risks, both of which vary by user and mini-program.

To formalize this trade-off, we decompose the cost for user i sharing data with mini-program j as:

$$c_{ij} = c_i + c_j + \epsilon_{ij},$$

where c_i represents the user's inherent privacy concerns, c_j reflects the mini-program's privacy risk (e.g., sensitivity of required data or reputation), and ϵ_{ij} is an idiosyncratic noise term.

Similarly, the benefit for user i from mini-program j is expressed as:

$$b_{ij} = b_i + b_j + \varepsilon_{ij},$$

where b_i captures the user's general receptiveness to digital services, b_j represents the utility offered by the mini-program, and ε_{ij} is an independent noise term.

A user authorizes data sharing if the benefit exceeds the cost:

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

Controlling for mini-program-specific factors, the decision to share data is primarily driven by the user's intrinsic trade-off, $b_i - c_i$.

We begin with a baseline scenario where b_i and c_i are unrelated, meaning a user's valuation of digital services is independent of their privacy concerns. This assumption aligns with common discussions on data privacy, where privacy apprehensions are treated separately from the demand for digital services. Based on this perspective, we posit the following hypothesis:

Hypothesis 1: *Holding all else constant, users with stronger privacy concerns authorize data sharing with fewer mini-programs.*

This hypothesis aligns with the conventional understanding of the relationship between data privacy and data sharing behavior. However, the pattern observed in Figure 2 contradicts this hypothesis. To account for differences among users beyond privacy concerns, we formally test this hypothesis using a user-month level panel regression approach that controls for various user characteristics:

$$Y_{it} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \lambda_t + \epsilon_i\,, \quad (3)$$

where the dependent variable Y_{it} represents user i's certain behavior (either the number of data-sharing authorizations or initial visits to mini-programs) in month t, Age_i and $Digital\ Experience_i$ serve as control variables, δ_i represents fixed effects related to other user characteristics, including gender and city; and λ_t represents time fixed effects.

Panel A of Table 4 reports the regression results. Columns (1) and (2), which use the number of data-sharing authorizations in a month as the dependent variable, show that the coefficients a_1 and a_2 are both insignificant, with or without controls. This finding reveals that "*concerned*" and "*very concerned*" users authorize data sharing at the same level as "*unconcerned*" users. This result confirms the paradox suggested by Figure 2.

Moreover, Columns (3) and (4), which use the number of initial visits to mini-programs as the dependent variable, reveal that "concerned" and "very concerned" users visit significantly more mini-programs.

Since some users may not visit any mini-programs in a given month, the user-month sample includes a large number of observations with zero authorizations. Conditioning on users with nonzero visits substantially reduces the number of zero observations. Columns (5) and (6) present the regression results from this conditional sample. In contrast to Columns (1) and (2), "*concerned*" and "*very concerned*" users are found to have significantly fewer authorizations.

To further investigate, the analysis is extended to the user–mini-program level for all possible user–mini-program pairs. This approach allows for controlling mini-program characteristics and conditioning on initial visits, as specified in Equation (2). The dependent variable Y_{ij} equals 1 if user i authorizes data sharing with or initially visits mini-program j, and 0 otherwise.

Panel B of Table 4 presents the results. Across all possible user–mini-program pairs, Columns (1) and (2) indicate that "concerned" and "very concerned" users do not significantly differ from "unconcerned" users in their likelihood of authorizing data sharing, regardless of whether mini-program fixed effects are included. Columns (3) and (4) reveal that "concerned" and "very concerned" users are significantly more likely to initially visit a given mini-program, even after controlling for mini-program fixed effects. Restricting the analysis to user–mini-program pairs with initial visits, columns (5) and (6) show that "concerned" and "very concerned" users are significantly less likely to authorize data sharing, irrespective of whether mini-program fixed effects are included.[10]

The contrast between the unconditional likelihood of authorizing data sharing and the likelihood conditional on initial visits highlights the nuanced nature of the data privacy paradox. While "concerned" and "very concerned" users appear to engage with more mini-programs overall, their lower likelihood of authorizing data sharing after an initial visit suggests a more cautious approach to privacy decisions.[11]

Table 5 extends the baseline analysis by comparing behavior during the post-survey period (August 2020 – December 2021) with the pre-survey period (July 2019 – July 2020). In Panel A, user-month regressions in columns (1) to (4) show that "concerned" and "very concerned" users authorized significantly more mini-programs and explored a wider variety of services after the survey. However, when the sample is restricted to months in which a user made at least one visit (columns (5) and (6)), the incremental effect on authorizations nearly disappears for "concerned" users and remains modest for "very concerned" users. This suggests that the intensified data privacy paradox among the "very concerned" is driven more by a surge in visiting activity than by a higher propensity to share data once a mini-program is visited.

[10] Table B4 extends the analysis by adding specific privacy concern types as controls. Consistent with Table 4, the coefficients for *Concerned* and *Very Concerned* remain statistically similar after accounting for data leakage and security (Type 1), price discrimination (Type 2), and seductive advertising (Type 3). The results also show that privacy concern types are not behaviorally equivalent. In the user-month analysis (Panel A), concerns about price discrimination and seductive advertising are positively and significantly associated with both mini-program visits and authorizations, even when conditioning on visits. A similar pattern emerges in the user–mini-program analysis (Panel B). These findings highlight that certain privacy concerns paradoxically correlate with greater engagement rather than caution, underscoring the need to distinguish between concern types when interpreting user behavior.

[11] Note that each user's consideration set is endogenous, as evidenced by the positive correlation between privacy concerns and the size of the consideration set. Consequently, while conditioning on each user's consideration set may improve estimation reliability by reducing the number of zero observations in the estimation sample, this comes at the cost of introducing an additional effect through the size of the consideration set, which may downwardly bias the estimated effect of privacy concerns on data-sharing authorization.

Panel B, which focuses on the user–mini-program level, reveals a similar pattern. While "concerned" users appear more likely to grant authorization after the survey across all user–mini-program pairs, this tendency diminishes when considering only pairs where at least one visit occurred.

Together with Table 4, these findings confirm the presence of the data privacy paradox in Alipay users' data sharing with third-party mini-programs while highlighting a key nuance: the paradox arises only in the unconditional relationship, where the number of authorized mini-programs is used as the outcome without accounting for the number of initially opened mini-programs. Notably, prior studies on the privacy paradox—e.g., Spiekermann, Grossklags, and Berendt (2001), Gross and Acquisti (2005), Norberg, Horne, and Horne (2007), and Athey, Catalini, and Tucker (2017)—are largely based on lab experiments and surveys. Our findings extend this analysis to a real-world context on a major financial platform and uncover richer nuances not previously documented.

C. The Role of Omitted Variables

To systematically explore the data privacy paradox, it is crucial to consider the possibility that users who are more apprehensive about sharing personal data may simultaneously perceive greater benefits from using mini-programs. If this assumption holds, the simple framework discussed in the previous subsection suggests that regressions of data-sharing authorization on privacy concerns may be misspecified. Such regressions fail to account for a user's underlying demand for digital services, which could be a key confounding factor.

To address this, we extend the use-month regression specified in equation (3) by incorporating measures of the user's digital demand:

$$Y_{it} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + a_5 Demand_{it} + \delta_i + \lambda_t + \epsilon_i, \quad (4)$$

where $Demand_{it}$ captures the user's demand for digital services. All other variables remain the same as specified in equation (1). We proxy this demand using the user's engagement with mini-programs on Alipay, under the assumption that more-active mini-program users derive greater benefits from these services. Specifically, we employ three measures: (1) the number of pages a

user opened in mini-programs within a month, (2) the number of launches, and (3) the number of sessions in which the user engaged with mini-programs within a month. Note that a session may include multiple launches. Due to the availability of the usage data, this analysis is restricted to the pre-survey period (July 2019 – July 2020).

Table 6 presents the regression results, revealing a consistent pattern across all specifications and measures of mini-program usage. Panel A reports results at the user-month level. Notably, the number of authorized mini-programs is positively and significantly correlated with mini-program usage, confirming that data-sharing decisions are closely tied to digital demand.

More importantly, once we control for digital demand, the relationship between data-sharing authorizations and privacy concerns reverses. The number of authorizations is now negatively related to both dummy variables, $Concerned_i$ and $Very\ Concerned_i$, with their coefficients significantly negative—in stark contrast to the positive coefficients reported in Panel A of Table 3. Furthermore, in each specification, the coefficient of $Very\ Concerned_i$ is more negative than that of $Concerned_i$, displaying a clear monotonic pattern—users with greater privacy concerns authorize data sharing with fewer mini-programs.

Panel B of Table 6 extends this analysis to the user–mini-program level, incorporating mini-program fixed effects alongside digital usage controls. The $Very\ Concerned_i$ dummy remains negative and significant, reinforcing the finding that users with greater privacy concerns authorize fewer mini-programs when digital demand is held constant.

These findings confirm that the apparent paradox—where privacy-concerned users share as much data as unconcerned users—largely arises from omitting digital demand from the analysis. Once differences in users' need or desire for digital services are properly accounted for, stronger privacy concerns indeed correspond to lower levels of data sharing, consistent with Hypothesis 1. This underscores the importance of further examining the relationship between digital demand and privacy concerns.

IV. Digital Engagement and Privacy Concerns

How does digital engagement influence users' privacy concerns and data sharing? Figure 1 illustrates a rising trend: as users gain more experience with digital services, their privacy concerns intensify. Previous research also suggests that concerns about data privacy are not entirely innate

but are shaped by experience—emerging as users engage with digital applications and share personal data (Stigler, 1980; Posner, 1981). The cost of privacy breaches, particularly when personal data is compromised (Fainmesser, Galeotti, and Momot, 2022), increases with the volume of shared data. Additionally, extensive data sharing enables digital providers to enhance price discrimination strategies (Taylor, 2004; Acquisti and Varian, 2005) and exploit user vulnerabilities (Liu et al., 2020). Consequently, as consumers share more data, their privacy concerns are likely to grow.

These mechanisms suggest that greater digital engagement amplifies privacy concerns, leading to the following hypothesis:

Hypothesis 2: *All else equal, more-active users of mini-programs are more likely to have stronger concerns about data sharing.*

In this section, we examine the causal relationship in two steps. In Subsection A, we employ an instrumental variable (IV) approach to demonstrate that increased digital usage leads to stronger privacy concerns, as reflected in the survey responses. In Subsection B, we analyze an incident that heightened Alipay users' awareness of data privacy, showing that active mini-program users became more prone to privacy concerns.

A. Causal Effect of Digital Usage on Privacy Concerns

The widespread adoption of Alipay in China is largely driven by its integration of various daily-life services, including a bike-sharing within its mini-program ecosystem. Between July 2017 and July 2020, China's bike-sharing sector grew rapidly, with companies like Ofo and Hellobike partnering with payment platforms like Alipay and WeChat Pay to enable bike access via digital wallets. As these companies expanded their fleets and coverage, local residents increasingly relied on Alipay, initially for bike sharing and later for other mini-programs. Since bike placements are determined by business strategies of third-party bike sharing companies rather than privacy concerns of users, they serve as a useful instrument to analyze how exogenously influenced digital engagement impacts privacy concerns.

Ouyang (2022) used a similar strategy to examine how mobile payment adoption, instrumented by the placement of Alipay-bundled shared bikes, facilitates consumer credit. His

analysis shows that bike placement follows a staggered, largely exogenous pattern across cities, driven by competitive pressures, logistical constraints, and capital availability.

We apply a two-stage least squares (2SLS) framework, with the IV analysis results presented in Table 7. In panel A, we use three measures of user mini-program engagement during the pre-survey period (July 2019 –July 2020): (1) the number of pages opened, (2) the number of launches, and (3) the number of sessions. These measures are log-transformed for the analysis.

Panel A2 reports the first-stage regressions of mini-program usage on the instrument, $log(Prior\ Bike\ Placement)_c$, which is the log-transformed average monthly number of Alipay-bundled shared bicycles placed in the user's city before the survey (July 2017–July 2020). Columns (1)–(2) use $log(1+\#Visited\ Pages)_{it}$ to measure engagement, columns (3)–(4) use $log(1+\#\ App\ Launches)_{it}$, and columns (5)–(6) rely on $log(1+\#\ App\ Uses)_{it}$. Controls include gender, education, job fixed effects, age, and digital experiences, with standard errors clustered at the city level to account for repeated observations within the same city over time. The results confirm that the instrument is strong, with F-statistics ranging from 26 to 51 across specifications, confirming it reliably predicts mini-program usage in the sample.

A valid instrument must satisfy the exclusion restriction, meaning shared bike placement should influence privacy concerns only through increased mini-program usage, with no direct effect. There is no direct link between bike placement and individuals' privacy concerns—seeing more bikes on the street does not inherently change attitudes toward data sharing with Alipay's mini-programs. Any effect arises solely from using shared bikes, which increases engagement with Alipay and its mini-program ecosystem.[12]

Panel A1 of Table 7 presents the second-stage regressions, linking instrumented mini-program usage to users' stated privacy concerns. The dependent variable is an indicator for whether a user reports being "*concerned*" or "*very concerned*" about data sharing with mini-programs. The endogenous regressor uses the fitted values of users' mini-program usage from the first-stage regressions in Panel A2.

The results consistently show a positive relationship between instrumented mini-program usage and privacy concerns. In columns (1)–(2), a 1% increase in the fitted number of visited pages

[12] One potential concern is that a user's prior experience with shared bikes could correlate with their privacy preferences. However, since data-sharing consent was required for prior shared bike usage, these users were likely less concerned about data privacy to begin with. This further reduces the likelihood that the instrument directly affects users' privacy concerns, thereby strengthening its validity.

raises the probability of expressing privacy concerns by about 0.078 percentage points. Columns (3)–(4) show a larger effect of 0.095 for app launches, while columns (5)–(6) report an effect size of 0.107 for app uses. These findings suggest that greater engagement with Alipay's ecosystem, even when exogenously driven, significantly amplifies data-sharing concerns. This effect is both statistically and economically significant, supporting Hypothesis 2 that deeper immersion in data-driven apps heightens anxieties about data privacy.

Panel A3 presents OLS results regressing privacy concerns on actual mini-program usage. The smaller OLS coefficients compared to the 2SLS estimates suggests that OLS likely underestimates the true effect due to selection bias, which the 2SLS approach addresses. For example, individuals less concerned about privacy may engage more with mini-programs, making the observed correlation between usage and privacy concerns appear weaker. In other words, low-privacy-concern individuals may already be "high-usage" users and remain less likely to report privacy worries, attenuating the relationship. This highlights the importance of instrumental variables for identifying causal effects in settings where observational data may be confounded by selection bias.

When comparing specifications with and without control variables, the gap between 2SLS and OLS coefficients narrows: 2SLS estimates decrease, while OLS estimates increase. This suggests that including controls accounts for observable confounders, improving the reliability of OLS estimates. However, the 2SLS approach remains crucial for addressing unobservable biases, reinforcing the robustness of the results.[13]

In panel B of Table 7, we further examine how mini-program engagement influences specific privacy concerns: data leakage and security (*Concern Type 1*) in columns (1)–(2), price discrimination (*Concern Type 2*) in columns (3)–(4), and seductive advertising (*Concern Type 3*) in columns (5)–(6). All regressions cluster at the city level, with selected specifications controlling for gender, education, job, age, and digital experience.

[13] Table B5 presents a robustness check on the causal link between digital usage and privacy concerns, using an alternative IV: *Prior Bike Placement Per Capita*—the average number of Alipay-bundled shared bikes in a city, normalized by the number of active Alipay users prior to July 2020—instead of the log transformation of raw bike count. This normalization accounts for city-level differences in city size and digital penetration. The table mirrors the main specification in Panel A of Table 7. Panel 1 shows that instrumented digital usage remains strongly and significantly correlated with privacy concerns, with effect sizes similar to the core analysis. Panel 2 confirms that per-capita bike placement is a strong first-stage instrument. These results reinforce that exogenous increases in digital engagement, driven by shared bike rollout, causally heighten users' privacy concerns across both main and alternative specifications.

The 2SLS results in Panel B1 show that increased engagement modestly but significantly raises concerns about data leakage in column (1), though the effect becomes smaller and insignificant with additional controls in column (2). For price discrimination, higher engagement is associated with increased concerns, with coefficients significant at the 10% level in columns (3)–(4). In contrast, the relationship with seductive advertising is negative and insignificant in columns (5)–(6), indicating no clear impact.

In summary, greater engagement with mini-programs is linked to heightened concerns about data leakage and price discrimination but not seductive advertising. These associations are modest and vary in strength depending on the model specification and controls.

B. Heterogeneous Responses to a Privacy Related Incident

To further examine how privacy concerns grow among users with different levels of digital engagement, we now analyze a notable event. On January 3, 2018, Alipay released its Annual User Footprint Report on the platform, summarizing users' Alipay activities in 2017. By default, the report included a pre-checked box agreeing to the *Sesame Credit Service Agreement*, which could enroll users in Alipay's credit scoring service without explicit consent if they overlooked the setting. This design flaw was quickly exposed and widely criticized on Chinese social media, leading to intense public scrutiny.

In response, Alipay disabled the default option the same day and issued a public apology, assuring users that those who had inadvertently consented would not be enrolled in the Sesame Credit service. However, the incident significantly heightened public awareness of data privacy issues, prompting a notable increase in Alipay users revoking data-sharing authorizations with mini-programs, as shown in Figure B6. This event serves as a natural experiment, enabling us to assess how users with varying levels of digital engagement reacted to heightened privacy concerns triggered by the incident.

We specifically examine whether heavy users of mini-programs exhibited stronger reactions, potentially indicating heightened privacy concerns, as posited by Hypothesis 2.

To test this hypothesis, we adopt an event study framework and estimate the following regression:

$$Daily\ Cancellation\ Dummy_{i,t} = \alpha_0 + \sum_{\substack{\tau=-5, \\ \tau \neq -1}}^{5} \beta_{H,\tau} \cdot Heavy\ User_i \cdot \mathbb{I}(t = \tau)$$

$$+ \beta_{H,6} \cdot Heavy\ User_i \cdot \mathbb{I}(t \geq 6) + \sum_{\substack{\tau=-5, \\ \tau \neq -1}}^{5} \beta_{L,\tau} \cdot Light\ User_i \cdot \mathbb{I}(t = \tau)$$

$$+ \beta_{L,6} \cdot Light\ User_i \cdot \mathbb{I}(t \geq 6) + \delta_i + \varepsilon_{i,t}, \quad (5)$$

where t is the number of days after the incident, $Daily\ Cancellation\ Dummy_{i,t}$ is a binary variable indicating whether user i canceled at least one mini-program on day t, $Heavy\ User_i$ is a dummy equal to 1 if user i was among the top 25% of mini-program users in the sample as of November 30, 2017, $Light\ User_i$ is defined as $1 - Heavy\ User_i$. The term δ_i represents individual fixed effects.

Since this event predates our survey sample, we use a different dataset of 100,000 users randomly selected from all active Alipay users. Details of this random sample are provided in the Appendix, where it is also used to validate the key findings established from our survey sample. Using this random sample, we estimate the regression specified above. Panel A of Figure 3 presents the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients. Consistent with Hypothesis 2, heavy users of mini-programs exhibit significantly stronger responses to the incident, as evidenced by a greater likelihood of canceling data-sharing authorizations. However, this response was temporary, possibly due to Alipay's swift corrective actions and the gradual fading of the incident from social media discussions. This finding holds when testing the differential response between heavy and light users, as shown in Panel A of Figure B7.

One might argue that heavy users' higher propensity to cancel data sharing reflects their better familiarity with Alipay's authorization settings rather than heightened privacy concerns triggered by the incident. To address this, we focus on a subsample of users from the random sample who had canceled data sharing with at least one mini-program before November 30, 2017. This filter ensures that all users in the subsample were already knowledgeable about how to cancel data-sharing prior to the incident.

Panel B of Figure 4 presents the estimated $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients for this subsample. While the behavioral gap between heavy and light users narrows, it remains significant, with heavy users still more likely to cancel data sharing. This reduced gap suggests that familiarity with the app plays a role but does not fully explain the higher cancellation rates among heavy users. For this subsample, the differential response between heavy and light users is statistically significant on days 0, 2, and 3 following the incident, as shown in Panel B of Figure B7.

These results are robust to alternative definitions of heavy users. In Panels A and B of Figure B8, we redefine heavy users as those in the top 50% of mini-program usage instead of the top 25%. The results remain consistent, with heavy users still exhibiting significantly stronger responses than light users. This confirms that our main findings do not sensitive to the specific cutoff used to classify user types.

Taken together, our analysis of Alipay users' responses to the privacy-related incident on January 3, 2018, supports Hypothesis 2 and confirms that users with greater digital engagement become more concerned about data privacy following the incident.[14] These findings reinforce the notion that privacy concerns grow with increased use of digital services.

V. Data Sharing and Credit Lines

In the growing digital economy, data sharing has become an essential tool for financial service providers to assess consumer creditworthiness and deliver more convenient financial services. As FinTech and BigTech lenders integrate alternative data into their models, data sharing increasingly shapes consumer credit decisions.

Berg et al. (2020) demonstrate the predictive power of digital footprints in assessing default risk and expanding credit access, showing how nontraditional user data enhances lending decisions. Similarly, Huang et al. (2020) and Gambacorta et al. (2023) use Alipay data to show that digital footprints significantly improve credit risk assessment. Building on these findings, we test the following hypothesis:

Hypothesis 3: *More data sharing improves users' credit access.*

This section tests this hypothesis by analyzing how data sharing with mini-programs on Alipay correlates with users' credit lines. The analysis uses a representative sample from Ouyang (2022), which includes users' credit access data. The summary statistics are shown in Table B6. The dataset comprises panel data on 41,485 randomly selected Alipay users observed over a 41-month

[14] This analysis complements the work of Agarwal et al. (2024), who also examined the adverse effects of data breaches on digital payment platforms, focusing on a different aspect: a privacy-related incident that heightened data privacy awareness without involving actual data breach. We use detailed data on individual usage of mini-programs and instances where users revoked data-sharing authorizations to closely examine personal behavior and differences among users based on their digital engagement. Although our context differs, our findings echo those of Agarwal et al. (2024) in that we observe short-lived reactions from consumers and a prevailing preference for convenience over privacy concerns over time.

period (May 2017 –September 2020). Each user appears in an average of 32 months of observations. The data integrates static user characteristics with time-varying measures, such as in-person payment flows, data-sharing decisions, and credit access. The average user was born in 1983, with 54% male and 88% lacking a bachelor's degree or higher.

Table 8 employs Poisson regressions to model the relationship between data sharing and credit lines, using a nonnegative dependent variable with many zeros. Columns (1), (3), and (5) represent different time horizons—t, t+1, and t+3—while columns (2), (4), and (6) correspond to the same periods but include a different set of controls. The analysis includes the following key variables:

- $log(1 + \#\ Authorized\ Mini\text{-}Programs)_{it}$: the number of mini-programs a user has authorized to access their data.
- $log(1 + \#\ Canceled\ Mini\text{-}Programs)_{it}$: a proxy for data-sharing withdrawals.
- $log(1 + In\text{-}Person\ Digital\ Payment\ Amount)_{it}$: a control variable capturing monthly in-person Alipay transaction volume (used in columns (1), (3), and (5)).

Including the last variable accounts for the relationship between credit lines and payment flows, as documented in Ouyang (2022). Additionally, columns (2), (4), and (6) control for digital payment amounts more precisely by including absorbed dummies for each centile of the variable.

The results show that greater digital payment flows are associated with higher credit limits, consistent with Berg et al. (2020), which found that richer digital footprints lead to more favorable lending terms. More importantly, a higher cumulative count of authorized mini-programs positively correlates with credit lines, indicating that consistent data sharing reduces informational asymmetry and signals user creditworthiness.

In practice, BigTech platforms like Alipay frequently update credit lines—often weekly or even daily—allowing newly granted data-sharing permissions to quickly feed into risk models and potentially increase credit limits. This immediate effect persists over time, as continuous data streams refine the platform's assessment of user creditworthiness.

Conversely, the cumulative number of canceled mini-programs shows negative coefficients, suggesting that frequent withdrawals of data-sharing permissions may raise lender concerns about transparency or stability. Although these effects weaken over longer horizons, they remain statistically and economically significant up to month t+3. This pattern reflects how revoking data-sharing permissions halts new information flow, reducing the platform's confidence in updating risk assessments.

Following the parametric analysis in Table 8, Table B7 explores the same relationship between data sharing and credit access using a nonparametric, bin-based approach that flexibly captures nonlinear effects. By grouping the number of authorized and canceled mini-programs into mutually exclusive bins (e.g., 0, 1, 2, 3–5, 6–10, and 10+), the table highlights how the marginal impact of data sharing varies across intensity levels. The results show a clear pattern of diminishing returns to authorization: while users with just one or two mini-programs see significant boosts in credit lines, the gains taper off for those with 10 or more. Importantly, these effects remain strong and persistent across all horizons from t to t+6, which corresponds to six months after the data sharing decision. This sustained impact underscores the long-lasting influence of digital footprint expansion—data authorizations continue to inform the platform's credit assessments well after the initial action, with only slowly decaying effects.

On the flip side, cancellations of mini-program authorizations display a steadily intensifying negative relationship with credit access. Users who cancel one or two programs experience only marginal or statistically weak penalties, but these effects become much more pronounced for users who revoke data access from five or more programs. The negative coefficients persist across the entire six-period horizon, again indicating a slow decay in their effect over time. This asymmetry, with diminishing returns to sharing but increasing penalties for revocation, suggests that BigTech platforms might place high value not just on the volume of shared data but on the stability and continuity of that sharing. Revoking access appears to disrupt the flow of behavioral signals that inform credit models, leading to a reduction in lenders' confidence and, subsequently, more constrained credit lines.

To identify the causal effect of data sharing on credit access, Table 9 employs a 2SLS framework that leverages the panel structure of the data. This setup enables identification strategies that go beyond simple cross-sectional comparisons by utilizing within-user variation over time. Following Ouyang (2022), the primary instrument is $log(Bike\ Placement)_{c,t}$, which captures monthly fluctuations in bike placement within each user's city. This variable is used to reflect exogenous shifts in users' data-sharing behavior.

This approach improves upon the IV analysis from the prior section, which relied on average bike availability before the survey to analyze privacy concerns. By incorporating time-varying instruments, the analysis isolates contemporaneous shocks to digital engagement that evolve

alongside users' credit access. This enhances both causal identification and the policy relevance of the findings.

Table 9 presents the 2SLS estimates. Panel B confirms the strength and relevance of the instrument: high F-statistics indicate that *log(Bike Placement)*$_{c,t}$ and digital payment flows are strong predictors of users' data-sharing decisions. This relationship holds across both contemporaneous and lagged models.

Panel A shows a strong positive causal effect of the instrumented number of authorized mini-programs on three measures of credit access: *log(1+x)* transformed credit line, a binary indicator for credit access, and conditional log credit line. These effects are robust across both contemporaneous specifications in columns (1) and (3), as well as those including lags in columns (2) and (4).

Notably, the results reveal substantial increases in the probability of obtaining credit on the extensive margin. The estimated coefficients for credit access in columns (3) and (4) are both substantial, at approximately 1.8. However, on the intensive margin—conditional on already having credit access—the increase in credit line (columns (5) and (6)) is positive but not statistically significant. These findings suggest that data sharing primarily supports users' initial entry into the credit system rather than substantially increasing credit amounts for those already approved.

Panel C presents OLS estimates, which, while directionally consistent, tend to show larger effects on the intensive margin than the 2SLS estimates, suggesting possible upward bias from selection effects.

In summary, Table 9 supports Hypothesis 3 while refining its interpretation. Richer digital footprints from authorized data sharing causally increase users' likelihood of obtaining credit, rather than significantly expanding credit limits for those already eligible. This highlights the critical role of data sharing in reducing entry barriers to formal credit markets, particularly benefiting previously underserved users.

Taken together, our analysis shows that users who engage in expansive data sharing gain greater credit access, while those who frequently curtail or revoke data sharing may miss out on credit opportunities. These findings underscore the importance of data-sharing behavior in shaping digital footprints, which directly influence lenders' decisions and credit allocation.

VI. Conclusion

In this paper, we integrate survey and administrative data to examine how Alipay users' data-sharing behavior with third-party mini-programs relates to their privacy concerns, digital engagement and credit access. Our analysis uncovers three key findings: 1) users share personal data with third-party service providers despite privacy concerns; 2) privacy concerns grow with digital engagement; 3) data sharing enhances credit access.

The rising demand for data-driven, personalized financial services makes our study highly relevant for both financial institutions and regulators. Traditional financial regulation frameworks primarily focus on managing the risks faced by institutions arising from their balance sheets. However, the growing reliance on personal consumer data in delivering financial services makes the management of data security and privacy risks a critical component of risk management for these institutions. Our findings offer valuable insights into consumers' data-sharing decisions, highlighting that privacy concerns persist even when users have full autonomy to opt out.

Appendix on a Representative Sample

Our survey sample tends to include more-active users, as they are more likely to complete the survey. This bias raises concerns about the generalizability of our findings. To ensure robustness, we construct a representative sample of 100,000 users randomly drawn from all active Alipay users. The results from this sample are consistent with those from our survey sample, reinforcing the validity and applicability of our main findings. These robustness results are detailed in this Appendix. Additionally, we use this representative sample to examine the differential impact of a privacy awareness event on heavy versus light users of mini-programs, as reported in Section IV.B.

We constructed a random sample of 100,000 Alipay users, randomly selected from all active Alipay users. Summary statistics for this sample are reported in Table B8 of the Online Appendix. On average, users in this random sample are 36.6 years old and have 60.7 months of digital experience, indicating that they tend to be older and have a shorter digital experience compared to our survey sample. The number of visited and authorized mini-programs in the random sample is only about one-third of that in the survey sample. Additionally, 12% of users in the random sample canceled data sharing with at least one mini-program, compared to 48% in the survey sample. Regarding mini-program usage, the average values of the four usage measures in the random sample are less than half of those observed in the survey sample.

Since users in the random sample did not participate in our survey, we cannot rely on their survey responses to measure privacy concerns. Instead, we use *Privacy Setting Changed*$_i$, a binary indicator of whether a user has modified Alipay's default privacy settings, as a behavior-based measure of privacy concerns. This approach follows Gross and Acquisti (2005), who used changes to Facebook's default data-sharing settings as a key indicator of users' privacy concerns.[15]

In Table A1, we present results using this behavior-based measure to re-examine the three key findings in the random sample. Panel A revisits the correlation between users' privacy concerns and their initial visits to mini-programs or data-sharing authorizations, building on the regression model used in Table 4, Panel A. The results confirm that users with greater privacy concerns tend to authorize data sharing with a larger number of mini-programs. This relationship remains

[15] Relative to the survey-based measure, this behavior-based measure is more objective as it is immune to noise in the survey, but it is also affected by the user's knowledge about how to change Alipay's default privacy settings. Despite this potential weakness, we can still use this behavior-based measure, after suitable control for user knowledge, to examine how privacy concerns are related to data-sharing authorization and cancellation.

significant even after controlling for key indicators of user knowledge, such as digital experience and age, gender, and city-specific effects, as well as when conditioning on initially visiting mini-programs. These findings highlight a more pronounced data privacy paradox in the random sample.

Panel B examines the privacy paradox while controlling for mini-program usage, using a method similar to that in Panel A of Table 6. The results reaffirm that the paradox becomes less pronounced after accounting for different usage variables, as the coefficient on *Privacy Setting Changed*$_i$ is generally smaller than in Table A, where users' digital demand is not controlled for.

Panel C reexamines the causal relationship between mini-program usage and privacy concerns in the representative sample, following the approach used in Table 7. Once again, we find a significant positive causal relationship between mini-program usage and privacy setting changes, which serve as a proxy for privacy concerns. This analysis mirrors that of Table 7 and reinforces the causal link between digital usage and heightened privacy concerns.

Taken together, these results confirm the robustness of our main findings within a representative sample of Alipay users, further reinforcing the overall conclusions of our analysis.

References

Acquisti, A. (2004). Privacy in Electronic Commerce and the Economics of Immediate Gratification. *Proceedings of the 5th ACM Conference on Electronic Commerce,* 21–29.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age. *Journal of Consumer Psychology*, 30(4), 736–758.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What Is Privacy Worth? *Journal of Legal Studies*, 42(2), 249–274.

Acquisti, A., & Varian, H. R. (2005). Conditioning Prices on Purchase History. *Marketing Science*, 24(3), 367–381.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442–492.

Agarwal, S., Ghosh, P., Ruan, T., & Zhang, Y. (2024). Transient Customer Response to Data Breaches of Their Information. Management Science.

Armantier, O., Doerr, S., Frost, J., Fuster, A., & Shue, K. (2021). Whom Do Consumers Trust with Their Data? US Survey Evidence. Bank for International Settlements.

Armantier, O., Doerr, S., Frost, J., Fuster, A., & Shue, K. (2024). Nothing to Hide? Gender and Age Differences in Willingness to Share Data. *BIS Working Papers* No 1187.

Athey, S., Catalini, C., & Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk (Working Paper No. 23488). National Bureau of Economic Research.

Babina, T., Bahaj, S., Buchak, G., De Marco, F., Foulis, A., Gornall, W., Mazzola, F. and Yu, T. (2025). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. *Journal of Financial Economics*, 169, 103950.

Ben-Shahar, O. (2016). Privacy is the New Money, Thanks to Big Data. *Forbes*.

Berg, T., Burg, V., Gombovic´, A., and Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7):2845–2897.

Berg, T., Fuster, A., & Puri, M. (2022). FinTech Lending. *Annual Review of Financial Economics*, 14, 187-207.

Bergemann, D., & Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1), 44–95.

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data, *American Economic Review* 91, 67–72.

Bian, B., Ma, X., & Tang, H. (2021). The Supply and Demand for Data Privacy: Evidene from Mobile Apps. *SSRN Electronic Journal*.

Bian, B., Pagel, M., & Tang, H. (2023). Consumer Surveillance and Financial Fraud. *SSRN Electronic Journal*.

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks. *Journal of Financial Economics*, 130(3), 453–483.

Chen, L., Bolton, P., Holmström, B. R., Maskin, E., Pissarides, C. A., Spence, A. M., Sun, T., Sun, T., Xiong, W., Yang, L., Huang, Y., Li, Y., Luo, X., Ma, Y., Ouyang, S., & Zhu, F. (2021). Understanding Big Data: Data Calculus in the Digital Era. *Luohan Academy Report*.

Cong, W., Xie, D., & Zhang, L. (2021). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science*, 67(10), 6480-6492.

Cooper, J. C., & Wright, J. (2018). The Missing Role of Economics in FTC Privacy Policy. *The Cambridge Handbook of Consumer Privacy*, 465.

Fainmesser, I. P., Galeotti, A., & Momot, R. (2022). Digital Privacy. *Management Science* 69(6):3157-3173.

Fang, H., Qin, X., Wu, W., & Yu, T. (2020). Mutual Risk Sharing and Fintech: The Case of Xiang Hu Bao. SSRN Electronic Journal.

Farboodi, M., & Veldkamp, L. (2021). A Model of the Data Economy. Working Paper, MIT and Columbia.

Fuller, C.S. (2019). Is the Market for Digital Privacy a Failure? *Public Choice*, 180(3), 353–381.

Gambacorta, L., Huang, Y., Li, Z., Qiu, H., & Chen, S. (2023). Data versus Collateral. *Review of Finance* 27(2), 369-398.

Huang, Y., Zhang, L., Li, Z., Qiu, H., Sun, T., & Wang, X. (2020). Fintech Credit Risk Assessment for SMEs: Evidence from China, SSRN.

Goldfarb, A., & Tucker, C. (2019). Digital Economics. *Journal of Economic Literature*, 57(1), 3–43.

Gross, R., & Acquisti, A. (2005). Information Revelation and Privacy in Online Social Networks (The Facebook Case). 11.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9), 2819–2858.

Lin, T. (2022). Valuing Intrinsic and Instrumental Preferences for Privacy. *Marketing Science*, 41(4):663-681.

Liu, H., Peng, C., Xiong, W., & Xiong, W. (2022). Taming the Bias Zoo. *Journal of Financial Economics,* 143, 716–741.

Liu, Z., Sockin, M., & Xiong, W. (2020). Data Privacy and Consumer Vulnerabilities. Working Paper, Princeton.

Mo, H., & Ouyang, S. (2025). (Generative) AI in Financial Economics. Working Paper, Saïd Business School, University of Oxford.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors. *Journal of Consumer Affairs,* 41(1), 100–126.

Ouyang, S. (2022). Cashless Payment and Financial Inclusion. Working Paper, Princeton.

Posner, R. A. (1981). The Economics of Privacy. *American Economic Review*, 71(2), 405–409.

Ramadorai, T., Uettwiller, A., & Walther, A. (2024). Privacy Policies and Consumer Data Extraction: Evidence from U.S. Firms. *Review of Finance*, forthcoming.

Rossi, A. G., & Utkus, S. P. (2024). The Diversification and Welfare Effects of Robo-Advising. *Journal of Financial Economics*, 157, 103869.

Solove, D. J. (2021). The Myth of the Privacy Paradox. *George Washington Law Review*, 89, 1–51.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd Generation E-commerce: Privacy Preferences Versus Actual Behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, 38-47.

Stigler, G. J. (1980). An Introduction to Privacy in Economics and Politics. *Journal of Legal Studies*, 9(4), 623-644.

Tang, H. (2020). The Value of Privacy: Evidence from Online Borrowers. Working Paper, London School of Economics.

Taylor, C. (2004). Consumer Privacy and the Market for Customer Information. *RAND Journal of Economics* 35 (4), 631–50.

Figure 1: Digital Experience and Privacy Concerns

This figure depicts the fraction of users indicating that they are "concerned" or "very concerned" about negative impacts caused by information shared with mini-programs in Alipay, across groups with different digital experiences, measured by the length of time since a user registered on Alipay. For each group, we also show the 95% confidence band of the mean estimate.

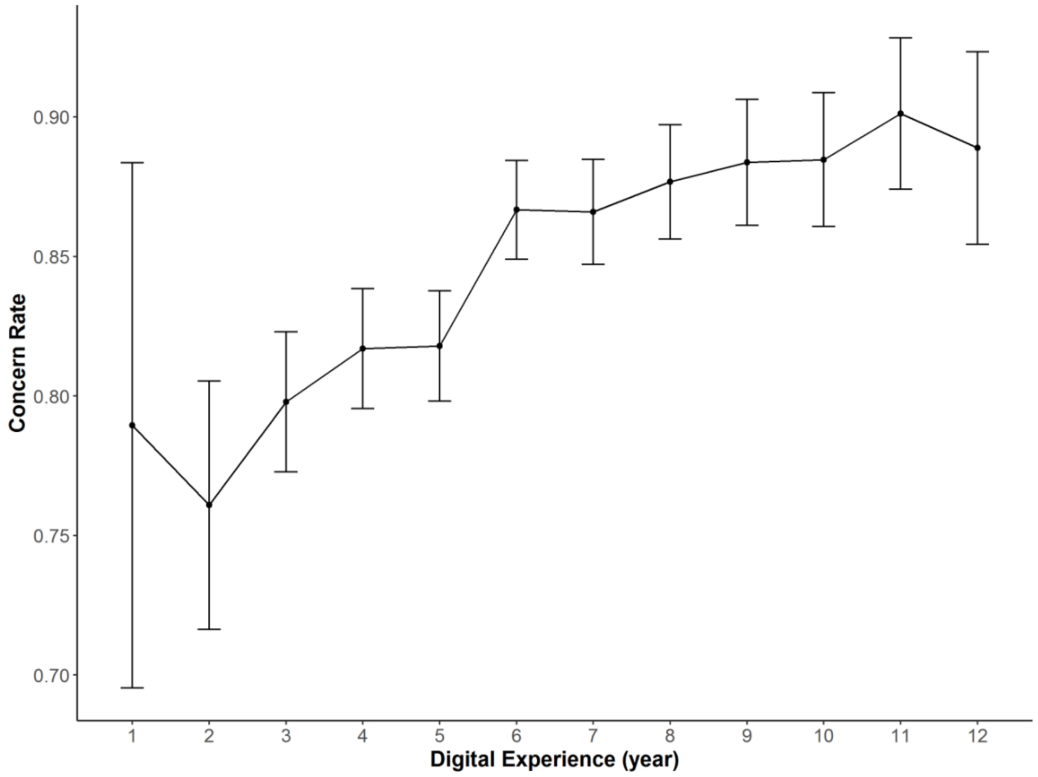


Figure 2: The Data Privacy Paradox

This figure depicts the numbers of data-sharing authorizations to mini-programs by Alipay users in three groups based on their answers to the question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" The pre-survey period is from July 2019 through July 2020, while the post-survey period is from August 2020 to December 2021.

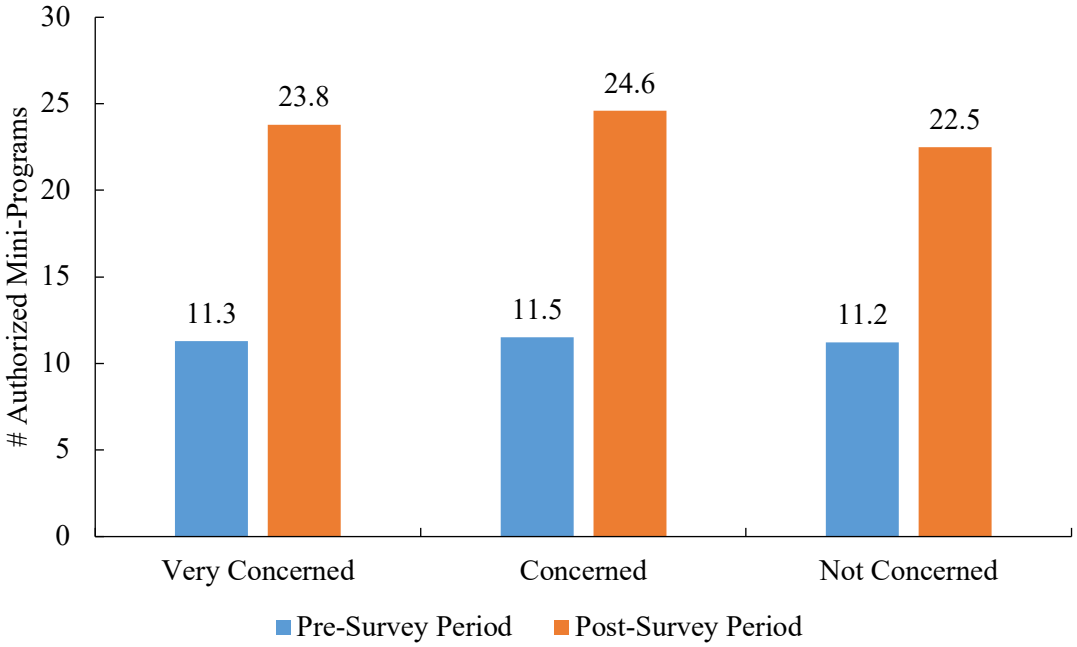
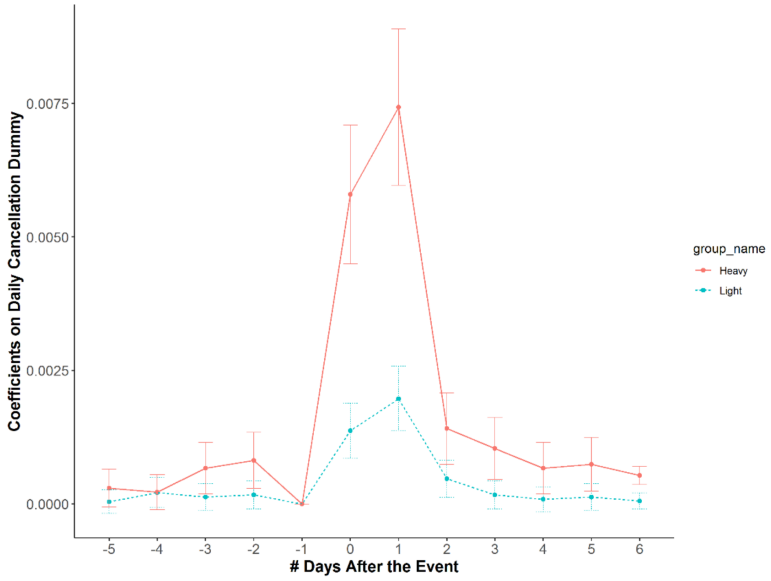


Figure 3: Activeness and Response to the 2017 Footprint Report Incident

The figures plot the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated by the regression specified in Equation (4), where the bands indicate 95% confidence intervals. Panel A covers the random sample of 100,000 Alipay users without any filtering, and Panel B covers only the users who had canceled data sharing with at least one mini-program before November 30, 2017, in the random sample. The data are at individual and daily levels. The sample period ranges from December 29, 2017 to January 31, 2018.

Panel A: Unfiltered Users



Panel B: Users with Cancellation before November 30, 2017

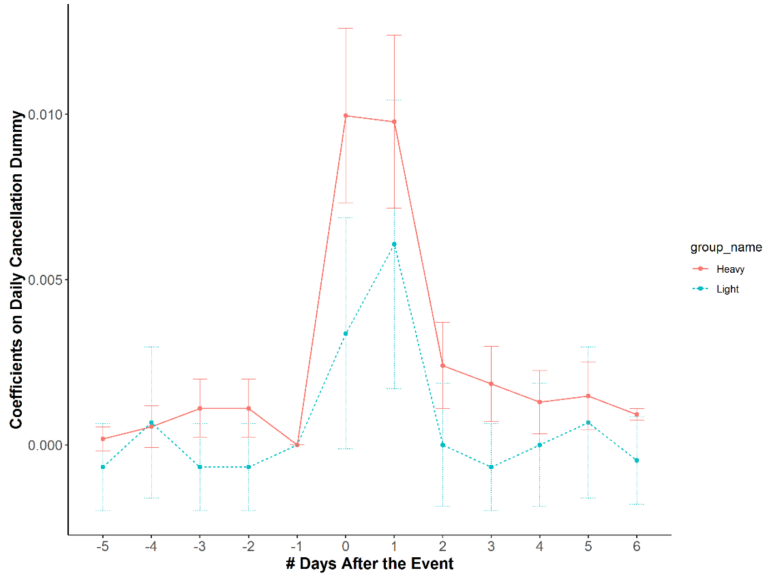


Table 1: Responses to Selected Survey Questions

This table summarizes responses to seven of the survey questions.

	Count	Total	Share
A. Are you concerned about privacy issues while using online services?			
Very concerned	13284	14250	93%
Concerned	882	14250	6%
Not concerned	84	14250	1%
B. What do you think about privacy protection in Alipay?			
Very good	6789	14250	48%
Ordinary	5600	14250	39%
Not good	679	14250	5%
No idea	1182	14250	8%
C. Do you know how to change privacy settings in Alipay?			
Yes	8529	14250	60%
No	5721	14250	40%
D. Have you ever changed your privacy settings in Alipay?			
Yes	5557	14250	39%
No	5025	14250	35%
No idea	3668	14250	26%
E. Have you ever used mini-programs in Alipay?			
Yes	10875	14250	76%
No	3375	14250	24%
F. 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%
G. What privacy issues are you concerned about when using mini-programs in Alipay? (multiple choice)			
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%

Table 2: Summary Statistics of the Survey Sample

This table reports summary statistics of the main sample of 10,875 users who finished the survey in July 2020 and indicated that they had used mini-programs in Alipay. Panel A reports user information in three parts. The first part reports the general information. *Concerned Dummy*$_i$ and *Very Concerned Dummy*$_i$ are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." *Privacy Setting Changed*$_i$, a proxy measure for privacy concerns, is a dummy variable equal to 1 if a user changed their privacy setting at least once between May 2017 and April 2020, and 0 otherwise. *Digital Experience*$_i$ is the number of months since the user first registered on Alipay, and *Age*$_i$ is the user's physical age in July 2020. The second part covers data sharing with mini programs, including the number of authorized and visited mini-programs over both the pre-survey period of July 2019 through July 2020 and the post-survey period of August 2020 through December 2021; the *Has Canceled*$_i$ status and # *Cancellations*$_i$, of used mini-programs over the pre-survey period of January 2013 to July 2020. The third part reports summary statistics of monthly use variables of Alipay users in each mini-program during the pre-survey period from July 2019 through July 2020, including the number of uses, the number of launches, and the number of visited pages. Use variables are winsorized at the 1% and 99% levels. Panel B reports the mean digital experience, age, female dummy, and education dummy for each group. *Female Dummy*$_i$ equals 1 if a user is female, and 0 otherwise. *Education Dummy*$_i$ equals 1 if a user has a college degree or higher, and 0 otherwise.

Panel A: User Information

	N	Mean	Std	Min	p25	Median	p75	Max
General Information								
Concerned Dummy$_i$	10,875	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Very Concerned Dummy$_i$	10,875	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Privacy Setting Changed$_i$	10,875	0.49	0.5	0.00	0.00	0.00	1.00	1.00
Digital Experience$_i$ (month)	10,871	74.97	35.07	4.00	48.00	70.00	97.00	190.00
Age$_i$ (year)	10,858	32.82	10.27	10.00	25.00	31.00	39.00	82.00
Data Sharing with Mini-Programs								
# Authorized Mini-Programs$_i$	10,875	34.22	22.78	0.00	19.00	30.00	43.00	422.00
# Visited Mini-Programs$_i$	10,875	46.57	55.45	1.00	26.00	38.00	53.00	1609.00
Has Canceled$_i$	10,875	0.48	0.50	0.00	0.00	0.00	1.00	1.00
# Cancellations$_i$	10,857	2.66	5.54	0.00	0.00	0.00	3.00	80.00
Monthly Mini-Program Use								
# App Uses$_{it}$	1,521,645	0.81	5.01	0.00	0.00	0.00	0.00	75.00
# App Launches$_{it}$	1,521,645	2.29	15.07	0.00	0.00	0.00	0.00	230.00
# Visited Pages$_{it}$	1,521,645	5.20	33.67	0.00	0.00	0.00	0.00	503.00

Panel B: Privacy Concern and Personal Characteristics

	Not Concerned$_i$	Concerned$_i$	Very Concerned$_i$	Difference	Difference
	(1)	(2)	(3)	(2) – (1)	(3) – (1)
Mean Digital Experience$_i$	66.868	75.725	76.961	8.857***	10.093***
				(1.018)	(0.996)
Mean Age$_i$	32.873	32.731	32.881	-0.142	0.008
				(0.300)	(0.293)
Mean Female Dummy$_i$	0.148	0.282	0.280	0.134***	0.132***
				(0.013)	(0.012)
Mean Education Dummy$_i$	0.137	0.221	0.214	0.084***	0.077***
				(0.012)	(0.012)

Table 3: Validating Survey-Based Privacy Concerns

This table reports how the survey-based measure of privacy concerns is related to privacy-seeking actions, including canceling data-sharing authorizations with mini-programs and changing Alipay's default privacy settings. *Concerned Dummy$_i$* and *Very Concerned Dummy$_i$* are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." Panel A shows results for user-level regressions. In columns (1)–(2), the dependent variable, *Has Canceled$_i$*, is a dummy that indicates whether a user has canceled at least one data-sharing authorization in the period of January 2013 through July 2020. In columns (3)–(4), the dependent variable, *Privacy Setting Changed$_i$*, is a dummy that indicates whether a user has changed Alipay's default privacy settings the period of May 2017 through April 2020. Panel B shows results for regressions at the user-mini–program level. In each pair of user-mini–program and existing data-sharing authorization, the dependent variable is a dummy that indicates whether the user canceled the authorization in January 2013 through July 2020. In panel B, we include more types of privacy risks from the survey—namely data leakage and security (*Concern Type 1$_i$*), price discrimination (*Concern Type 2$_i$*), and seductive advertising (*Concern Type 3$_i$*). We cluster the standard errors at the user level. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User-Level Analysis

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

Panel B: Analysis at the User-Mini–Program Level

	Canceled Dummy$_{ij}$		
	(1)	(2)	(3)
Concerned Dummy$_i$	0.004		0.002
	(0.003)		(0.003)
Very Concerned Dummy$_i$	0.011***		0.009**
	(0.003)		(0.003)
Concern Type 1$_i$		0.010***	0.007**
		(0.003)	(0.003)
Concern Type 2$_i$		0.005*	0.004
		(0.003)	(0.003)
Concern Type 3$_i$		0.004*	0.004*
		(0.002)	(0.002)
Digital Experience$_i$ ($\times 10^{-4}$)	1.218***	1.057***	1.052***
	(0.305)	(0.306)	(0.307)
Age$_i$ ($\times 10^{-4}$)	2.547**	3.169***	3.154***
	(1.141)	(1.163)	(1.166)
Mini-program, City, and Gender FE	Y	Y	Y
Cluster by Individual	Y	Y	Y
Observations	480,542	480,542	480,542
Adjusted R^2	0.107	0.107	0.107

Table 4: The Data Privacy Paradox

These tables present regression analysis of the data privacy paradox results for the pre-survey period from July 2019 through July 2020. *Concerned Dummy*$_i$ and *Very Concerned Dummy*$_i$ are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." *Education*$_i$ is a dummy indicating whether the user has a college degree or higher. *Self Control*$_i$ is a dummy indicating whether the user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the pre-survey period. Panel A displays results for user-month-level regressions. Columns (1)–(2) present results for the monthly number of authorized mini-programs in the full sample, columns (3)–(4) for the monthly number of initially visited mini-programs in the full sample, and columns (5)–(6) for the monthly number of authorized mini-programs in the subsample with nonzero monthly initial visits. # *Authorized Mini-Programs*$_{it}$ and # *Visited Mini-Programs*$_{it}$ are the number of mini-programs authorized and visited by individual *i* at time *t*, respectively. Panel B provides results for regressions at the user-mini–program level. For each user–mini-program pair, columns (1)–(2) report results for the dummy indicating whether the user allowed authorization, columns (3)–(4) for the dummy indicating whether the user visited at least once, and columns (5)–(6) for the dummy indicating whether the user allowed authorization, conditional on pairs where the user visited at least once. *Authorized Dummy*$_{ij}$ equals one if a user *i* authorized data sharing to a mini-program *j*, otherwise zero. *Visited Dummy*$_{ij}$ equals one if a user *i* visited a mini-programs at least once *j*, otherwise zero. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User-Month-Level Analysis

	# Authorized Mini-Programs$_{it}$		# Visited Mini-Programs$_{it}$		# Authorized Mini-Programs$_{it}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$	0.023	0.014	0.089***	0.088***	-0.036	-0.032
	(0.015)	(0.015)	(0.023)	(0.023)	(0.027)	(0.026)
Very Concerned Dummy$_i$	0.008	-0.001	0.142***	0.140***	-0.092***	-0.086***
	(0.015)	(0.015)	(0.024)	(0.024)	(0.027)	(0.026)
Digital Experience$_{it}$		0.001***		-0.0002		0.0003
		(0.0002)		(0.0003)		(0.0003)
Age$_{it}$		-0.003***		0.015***		-0.011***
		(0.001)		(0.001)		(0.001)
Constant	0.800***		1.023***		1.798***	
	(0.013)		(0.020)		(0.023)	
City, Gender, Time FE	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	152,110	151,886	152,110	151,886	70,988	70,892
Adjusted R^2	0.00002	0.005	0.001	0.013	0.0003	0.016

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$		Visited Dummy$_{ij}$		Authorized Dummy$_{ij}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$ ($\times 10^{-4}$)	0.862	0.386	2.897***	2.552***	-306.483***	-296.651***
	(0.745)	(0.735)	(0.848)	(0.836)	(56.372)	(47.329)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	0.028	-0.465	3.755***	3.340***	-595.478***	-540.905***
	(0.736)	(0.728)	(0.846)	(0.840)	(55.658)	(46.779)
Digital Experience$_i$		5.517***		3.806***		267.192***
		(0.800)		(0.960)		(52.984)
Age$_i$		-1.958***		2.045***		-524.245***
		(0.287)		(0.367)		(19.010)
Constant	0.004***		0.005***		0.851***	
	(0.0001)		(0.0001)		(0.005)	
Mini-program, City and Gender FE	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	25,414,875	25,364,288	25,414,875	25,364,288	132,924	132,713
Adjusted R^2	0.000	0.105	0.000	0.129	0.003	0.148

Table 5: The Data Privacy Paradox in the Post-Survey Period

These tables present regression analysis of the data privacy paradox results comparing the post-survey period from August 2020 through December 2021 with the pre-survey period from July 2019 to July 2020. *Concerned Dummy*$_i$ and *Very Concerned Dummy*$_i$ are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." Panel A displays results for user-month-level regressions. Columns (1)–(2) present results for the monthly number of authorized mini-programs in the full sample, columns (3)–(4) for the monthly number of initially visited mini-programs in the full sample, and columns (5)–(6) for the monthly number of authorized mini-programs in the subsample with nonzero monthly initial visits. *# Authorized Mini-Programs*$_{it}$ and *# Visited Mini-Programs*$_{it}$ are the number of mini-programs authorized and visited by individual *i* at time *t*, respectively. I Panel B provides results for regressions at the user-mini–program level. For each user–mini-program pair, columns (1)–(2) report results for the dummy indicating whether the user allowed authorization, columns (3)–(4) for the dummy indicating whether the user visited at least once, and columns (5)–(6) for the dummy indicating whether the user allowed authorization, conditional on pairs where the user visited at least once. *Authorized Dummy*$_{ij}$ equals one if a user *i* authorized data sharing to a mini-program *j*, otherwise zero. *Visited Dummy*$_{ij}$ equals one if a user *i* visited a mini-programs at least once *j*, otherwise zero. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User-Month-Level Analysis

	# Authorized Mini-Programs$_{it}$		# Visited Mini-Programs$_{it}$		# Authorized Mini-Programs$_{it}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$ × Post-Survey Dummy$_i$	0.085***	0.085***	0.174***	0.174***	-0.005	-0.005
	(0.026)	(0.026)	(0.054)	(0.054)	(0.009)	(0.009)
Very Concerned Dummy$_i$ × Post-Survey Dummy$_i$	0.062**	0.062**	0.147***	0.143***	0.019**	0.018**
	(0.026)	(0.026)	(0.054)	(0.054)	(0.009)	(0.009)
Concerned Dummy$_i$	0.025	-0.002	0.097***	0.067**	-0.028***	-0.032***
	(0.016)	(0.017)	(0.025)	(0.027)	(0.008)	(0.008)
Very Concerned Dummy$_i$	0.010	-0.016	0.153***	0.135***	-0.069***	-0.071***
	(0.016)	(0.017)	(0.025)	(0.028)	(0.008)	(0.008)
Post-Survey Dummy$_i$	0.334***		0.371***		0.015**	
	(0.022)		(0.039)		(0.008)	
Constant	0.860***		1.101***		0.818***	
	(0.014)		(0.021)		(0.007)	
City, Gender, Time FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	337,032	336,505	337,032	336,505	75,510	75,418
Adjusted R^2	0.010	0.060	0.005	0.036	0.003	0.047

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$		Visited Dummy$_{ij}$		Authorized Dummy$_{ij}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$ × Post-Survey Dummy$_i$ (× 10^{-4})	274.587***	256.996***	3.314***	-9.945	-13.321	17.589
	(53.601)	(52.630)	(0.530)	(9.511)	(52.162)	(46.641)
Very Concerned Dummy$_i$ × Post-Survey Dummy$_i$ (× 10^{-4})	535.489***	502.886***	2.696***	-23.389**	84.511	94.071**
	(53.176)	(52.183)	(0.524)	(9.492)	(51.559)	(46.618)
Concerned Dummy$_i$ (× 10^{-4})	-272.477***	-255.647***	-0.0001	12.561	-272.477***	-266.498***
	(53.697)	(52.690)	(0.0001)	(9.630)	(53.697)	(46.427)
Very Concerned Dummy$_i$ (× 10^{-4})	-534.387***	-502.433***	-0.0001	25.502***	-534.387***	-488.048***
	(53.268)	(52.240)	(0.0002)	(9.620)	(53.269)	(45.895)
Post-Survey Dummy$_i$	-0.863***	-0.799***	-0.997***	-0.923***	0.026***	-0.030***
	(0.004)	(0.004)	(0.000)	(0.001)	(0.004)	(0.004)
Constant	0.866***		1.000***		0.866***	
	(0.004)		(0.000)		(0.004)	
Mini-program, City and Gender FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	72,525,375	72,401,144	72,525,375	72,401,144	330,483	329,967
Adjusted R^2	0.298	0.367	0.368	0.440	0.004	0.163

Table 6: Resolving the Data Privacy Paradox by Controlling Digital Usage

This table shows that after adding digital usage as controls, the data privacy paradox disappeared. *Concerned Dummy*$_i$ and *Very Concerned Dummy*$_i$ are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." In panel A, we use monthly number of authorized mini-programs from July 2019 through July 2020 as the dependent variable. We control digital usages in the regression, namely the number of visited pages in columns (1)–(2), the number of mini-programs launches columns (3)–(4), and the number of mini-programs uses columns (5)–(6). Panel B provides results for regressions at the user-mini–program level after controlling usage variables and mini-program fixed effects. *Authorized Dummy*$_{ij}$ equals one if a user *i* authorized data sharing to a mini-program *j*, otherwise zero. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We cluster the standard errors at the user level and report standard errors in parentheses.

Panel A: User-Month-Level Analysis

	# Authorized Mini-Programs$_{it}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$	-0.046***	-0.044***	-0.047***	-0.048***	-0.033***	-0.039***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Very Concerned Dummy$_i$	-0.062***	-0.057***	-0.063***	-0.062***	-0.053***	-0.057***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Digital Experience$_i$		0.00000		0.0002		0.0002*
		(0.0001)		(0.0001)		(0.0001)
Age$_i$		-0.006***		-0.008***		-0.010***
		(0.001)		(0.001)		(0.001)
log(1+# Visited Pages) $_{it}$	0.373***	0.377***				
	(0.004)	(0.004)				
log(1+# App Launches) $_{it}$			0.429***	0.441***		
			(0.005)	(0.005)		
log(1+# App Uses) $_{it}$					0.463***	0.497***
					(0.007)	(0.007)
Constant	0.082***		0.162***		0.361***	
	(0.011)		(0.011)		(0.011)	
City, Gender, and Time FE	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Observations	152,110	151,886	152,110	151,886	152,110	151,886
Adjusted R^2	0.228	0.232	0.204	0.212	0.146	0.158

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$ ($\times 10^{-4}$)	-0.344	-0.434	-0.330	-0.455	-0.228	-0.346
	(0.436)	(0.439)	(0.458)	(0.460)	(0.534)	(0.536)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	-0.839*	-0.950**	-0.931**	-1.075**	-1.095**	-1.243**
	(0.430)	(0.434)	(0.452)	(0.454)	(0.525)	(0.525)
log(1+# Visited Pages)$_{ij}$	0.285***	0.280***				
	(0.001)	(0.001)				
log(1+# App Launches)$_{ij}$			0.351***	0.344***		
			(0.001)	(0.001)		
log(1+# App Uses)$_{ij}$					0.427***	0.414***
					(0.002)	(0.002)
Constant	0.002***		0.002***		0.003***	
	(0.00004)		(0.00004)		(0.00004)	
Mini-program, City, and Gender FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Observations	25,414,875	25,364,288	25,414,875	25,364,288	25,414,875	25,364,288
Adjusted R^2	0.513	0.544	0.471	0.505	0.350	0.392

Table 7: Causal Effect of Digital Demand on Privacy Concerns

These tables examine the causal relationship between digital demand and revealed privacy concern in the survey. *log(Prior Bike Placement)*$_c$ is the log transformed average of the monthly number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* before the survey in July 2020. In Panel A, the dependent variable, *Concerned Dummy*$_i$, is a dummy variable that equals 1 if the user *i*'s answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." In Panel A, we use the number of visited pages in columns (1)–(2), the number of mini-programs launches in columns (3)–(4), and the number of mini-programs uses in columns (5)–(6) to capture the demand for digital services. All usage variables are calculated by summing the activities for each individual across all mini-programs during the pre-survey period from July 2019 through July 2020. Panel A1 outlines the 2SLS estimates; Panel A2 describes the first stage; Panel A3 provides the OLS results. In Panel B, we use three types of privacy risks from the survey as dependent variables—namely data leakage and security (*Concern Type 1*$_i$) in columns (1)–(2), price discrimination (*Concern Type 2*$_i$) in columns (3)–(4), and seductive advertising (*Concern Type 3*$_i$) in columns (5)–(6). In panel B, we use the number of visited pages to capture the demand for digital services. Panel B1 outlines the 2SLS estimates; Panel B2 describes the first stage; Panel B3 provides the OLS results. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: General Concerns

	Concerned Dummy$_i$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A1: Two-Stage Least Squares						
$\log(1 + \#\,\widehat{\text{Visited Pages}})_i$	0.121***	0.078***				
	(0.026)	(0.028)				
$\log(1 + \#\,\widehat{\text{App Launches}})_i$			0.145***	0.095***		
			(0.031)	(0.034)		
$\log(1 + \widehat{\#\,\text{App Uses}})_i$					0.158***	0.107***
					(0.037)	(0.039)
Panel A2: First Stage for Usage Variables						
log(Prior Bike Placement)$_c$	0.065***	0.069***	0.055***	0.056***	0.050***	0.050***
	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)
F-Statistic	50.04	25.59	37.66	34.92	25.65	50.99
Adjusted R^2	0.010	0.028	0.008	0.034	0.007	0.052
Panel A3: Ordinary Least Squares						
log(1+# Visited Pages)$_i$	0.013***	0.014***				
	(0.002)	(0.003)				
log(1+# App Launches)$_i$			0.014***	0.015***		
			(0.002)	(0.003)		
log(1+# App Uses)$_i$					0.015***	0.016***
					(0.002)	(0.003)
Adjusted R^2	0.003	0.034	0.003	0.034	0.004	0.034
Gender, Education, Job FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by City	Y	Y	Y	Y	Y	Y
Observations	10,871	6,785	10,871	6,785	10,871	6,785

Panel B: Specific Concern Types

	Concern Type 1$_i$		*Concern Type 2$_i$*		*Concern Type 3$_i$*	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B1. Two-Stage Least Squares						
$\log(1 + \#\ \widehat{\text{Visited Pages}})_i$	0.044***	0.033	0.056*	0.057*	-0.028	-0.016
	(0.023)	(0.023)	(0.033)	(0.032)	(0.036)	(0.048)
Panel B2. First Stage for Usage Variables						
log(Prior Bike Placement)$_c$	0.065***	0.069***	0.065***	0.069***	0.065***	0.069***
	(0.009)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
F-Statistic	50.04	25.59	50.04	25.59	50.04	25.59
Adjusted R^2	0.010	0.028	0.010	0.028	0.010	0.028
Panel B3. Ordinary Least Squares						
log(1+# Visited Pages)$_i$	0.004*	0.002	0.011***	0.011***	0.012***	0.010***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Adjusted R^2	0.0002	0.033	0.002	0.082	0.001	0.022
Gender, Education, Job FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by City	Y	Y	Y	Y	Y	Y
Observations	10,871	6,785	10,871	6,785	10,871	6,785

Table 8: Digital Footprints, Data Sharing, and Financial Credit Access

This table examines the relationship between mini-programs data sharing with financial credit access in Alipay using the user-month panel data in Ouyang (2022). We use Poisson Regressions in this table as there are many zeros in dependent variables. We use two measures of mini-programs data sharing. The first one, *log(1+# Authorized Mini-Programs)$_{it}$*, the log transformation of number of mini-programs authorized by individual *i* at time *t*, indicates how many data a user shared. The second one, *log(1+# Canceled Mini-Programs)$_{it}$,* is a proxy of withdraw from data sharing. Columns (1), (3), and (5) control for *log(1+In-Person Digital Payment Amount)$_{it}$*, the log transformation of in-person digital payment transactions, which is a key driver of credit lines in Alipay as shown by Ouyang (2022). Columns (2), (4), and (6) control for digital payment amount more precisely by including absorbed dummies for each centile of this variable. The results for regressions are at the user-month level. We use credit line in month *t*, *t+1*, and *t+3* as the dependent variables. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Credit Line$_{iT}$					
	t	t	t+1	t+1	t+3	t+3
	(1)	(2)	(3)	(4)	(5)	(6)
log(1+# Authorized Mini-Programs)$_{it}$	0.368***	0.387***	0.336***	0.356***	0.277***	0.296***
	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)
log(1+# Canceled Mini-Programs)$_{it}$	-0.181***	-0.165***	-0.177***	-0.161***	-0.160***	-0.145***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)
log(1+In-Person Digital Payment Amount)$_{it}$	0.007***		0.008***		0.009***	
	(0.001)		(0.001)		(0.001)	
Individual, Time FE	Y	Y	Y	Y	Y	Y
Digital Payment Amount Centile Dummies	N	Y	N	Y	N	Y
Cluster by Individual, Time	Y	Y	Y	Y	Y	Y
Observations	843,286	843,286	828,157	828,157	774,047	774,047
Adjusted Pseudo R^2	0.867	0.868	0.867	0.868	0.872	0.874

Table 9: Causal Effect of Data Sharing on Financial Credit Access

This table investigates the causal relationship between data sharing behaviors and individuals' access to financial credit in Alipay using the user-month panel data in Ouyang (2022). The dependent variables capture different aspects of credit access: $log(1+Credit\ Line)_{i,t}$ measures the logarithm of one plus the credit line amount for individual i at time t, $Has\ Credit_{i,t}$ is a dummy for any credit access, and $log(Credit\ Line)_{i,t}$ measures the log credit line amount conditional on having credit. $log(1 + \#\ Authorized\ Mini\text{-}Programs)_{it}$ is the log-transformed value of the number of mini-programs authorized by individual i at time t. $log(1+In\text{-}Person\ Digital\ Payment\ Amount)_{it}$, the log transformation of in-person digital payment transactions by individual i at time t, is a key driver of credit lines in Alipay as shown by Ouyang (2022). $log(Bike\ Placement)_{c,t}$ is the log transformed number of Alipay-bundled shared bicycles placed in the user i's city c at month t. Panel 1 outlines the 2SLS estimates; Panel 2 describes the first stage; Panel 3 provides the OLS results. Columns (1), (3), (5) use contemporaneous variables while columns (2), (4), (6) use lagged variables. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	$log(1+Credit\ Line)_{i,t}$		$Has\ Credit_{i,t}$		$log(Credit\ Line)_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Two-Stage Least Squares						
$\log(1 + \#\ \widehat{\text{Authorized Mini-Programs}})_{i,t}$	1.717***		0.175**		0.047	
	(0.530)		(0.068)		(0.261)	
$\log(1 + \#\ \widehat{\text{Authorized Mini-Programs}})_{i,t-1}$		1.782***		0.187**		0.106
		(0.573)		(0.074)		(0.251)
$\log(1+\text{In-Person Digital Payment Amount})_{i,t}$	0.199***		0.019**		0.113***	
	(0.067)		(0.008)		(0.029)	
$\log(1+\text{In-Person Digital Payment Amount})_{i,t-1}$		0.189**		0.017*		0.109***
		(0.074)		(0.009)		(0.029)
Panel B. First Stage for Data Sharing Variables						
$\log(\text{Bike Placement})_{c,t}$	0.015***		0.015***		0.013***	
	(0.004)		(0.004)		(0.004)	
$\log(\text{Bike Placement})_{c,t-1}$		0.014***		0.014***		0.013***
		(0.004)		(0.004)		(0.004)
$\log(1+\text{In-Person Digital Payment Amount})_{i,t}$	0.125***		0.125***		0.108***	
	(0.003)		(0.003)		(0.002)	
$\log(1+\text{In-Person Digital Payment Amount})_{i,t-1}$		0.128***		0.128***		0.111***
		(0.003)		(0.003)		(0.002)
F-Statistic	953.29	1020.2	953.29	1020.2	1162.9	1254.6
Adjusted R^2	0.295	0.299	0.295	0.299	0.262	0.268
Panel C. Ordinary Least Squares						
$\log(1+\#\ \text{Authorized Mini-Programs})_{i,t}$	1.615***		0.152***		0.552***	
	(0.062)		(0.007)		(0.024)	
$\log(1+\#\ \text{Authorized Mini-Programs})_{i,t-1}$		1.611***		0.152***		0.539***
		(0.063)		(0.008)		(0.025)
$\log(1+\text{In-Person Digital Payment Amount})_{i,t}$	0.212***		0.022***		0.058***	
	(0.009)		(0.001)		(0.003)	
$\log(1+\text{In-Person Digital Payment Amount})_{i,t-1}$		0.210***		0.021***		0.061***
		(0.010)		(0.001)		(0.003)
Adjusted R^2	0.318	0.293	0.235	0.205	0.244	0.244
City, Time FE	Y	Y	Y	Y	Y	Y
Cluster by City, Time	Y	Y	Y	Y	Y	Y
Observations	963,015	944,166	963,015	944,166	696,663	694,222

Table A1: Results from the Representative Sample

This table reports four sets of robustness tests from using the representative random sample of 100,000 Alipay users. Panel A presents the robustness test for the digital privacy paradox, where the regressions are at the user-month level. *Privacy Setting Changed* is a behavior-based measure for privacy concerns, defined as a dummy variable that equals 1 if a user changed the default privacy settings at least once between May 2017 and April 2020, and 0 otherwise. Columns (1)–(2) present results for the monthly number of authorized mini-programs in the full sample, columns (3)–(4) for the monthly number of initially visited mini-programs in the full sample, and columns (5)–(6) for the monthly number of authorized mini-programs in the subsample with nonzero monthly initial visits. In columns (2), (4) and (6), we control for digital experience and age, along with gender, city, and time fixed effects. Panel B tests the positive relationship between privacy concerns and demand for digital services, where the regressions are at the user-mini–program-month level, and the standard errors are clustered at the user level. We use monthly number of authorized mini-programs from July 2019 through July 2020 as the dependent variable. We control digital usages in the regression, namely the number of visited pages in columns (1)–(2), the number of mini-programs launches columns (3)–(4), and the number of mini-programs uses columns (5)–(6). Columns (1), (3), and (5) show regression results without any other controls, while columns (2), (4), and (6) control for digital experience and age, as well as user gender, user city, and time fixed effects. Panel C examines the causal relationship between digital demand and privacy setting, where the regressions are at the user-month level, and the standard errors are clustered at the city level. $log(Prior\ Bike\ Placement)_c$ is the log transformed average of the monthly number of Alipay-bundled shared bicycles placed in the user i's city c before the survey in July 2020. We use the number of visited pages columns (1)–(2), the number of mini-programs launches columns (3)–(4), and the number of mini-programs uses columns (5)–(6) to capture the demand for digital services. Panel C shows the Two-Stage Least Squares (2SLS) estimates, utilizing the interaction of city-level average bicycle placement and bikes using experience as an instrument for the digital demand, exploring the causal relationship between digital demand and privacy setting. Panel C1 outlines the 2SLS estimates; Panel C2 describes the first stage; Panel C3 provides the OLS results. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: Analysis at User-Month Level of the Data Privacy Paradox

	# Authorized Mini-Programs$_{it}$		*# Visited Mini-Programs$_{it}$*		*# Authorized Mini-Programs$_{it}$*	
	(1)	(2)	(3)	(4)	(5)	(6)
Privacy Setting Changed$_i$	0.158***	0.136***	0.020***	0.175***	0.157***	0.145***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.013)	(0.013)
Constant	0.195***		0.239***		1.333***	
	(0.001)		(0.001)		(0.003)	
City, Gender, Time FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	1,776,150	1,738,656	1,776,150	1,738,656	273,376	272,178
Adjusted R^2	0.004	0.020	0.004	0.018	0.002	0.018

Panel B: Analysis at User-Month Level of Privacy Concerns and Digital Demand

	# *Authorized Mini-Programs*$_{it}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Privacy Setting Changed$_i$	0.082***	0.071***	0.085***	0.074***	0.097***	0.083***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
log(1+# Visited Pages)$_{it}$	0.283***	0.278***				
	(0.001)	(0.001)				
log(1+# App Launches)$_{it}$			0.372***	0.364***		
			(0.002)	(0.002)		
log(1+# App Uses)$_{it}$					0.512***	0.499***
					(0.004)	(0.004)
Constant	0.092***		0.098***		0.120***	
	(0.001)		(0.001)		(0.001)	
City, Gender, and Time FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Observations	1,776,150	1,738,656	1,776,150	1,738,656	1,776,150	1,738,656
Adjusted R^2	0.179	0.187	0.169	0.176	0.135	0.142

Panel C: Two-Stage Least Squares

	Privacy Setting Changed$_i$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel C1: Two-Stage Least Squares						
$\log(1 + \#\ \widehat{\text{Visited Pages}})_i$	0.027***	0.023***				
	(0.005)	(0.005)				
$\log(1 + \#\ \widehat{\text{App Launches}})_i$			0.033***	0.028***		
			(0.006)	(0.007)		
$\log(1 + \widehat{\#\ \text{App Uses}})_i$					0.046***	0.038***
					(0.009)	(0.010)
Panel C2: First Stage for Usage Variables						
$\log(\text{Prior Bike Placement})_c$	0.116***	0.113***	0.095***	0.093***	0.069***	0.067***
	(0.017)	(0.018)	(0.015)	(0.016)	(0.013)	(0.013)
F-Statistic	44.40	147.68	38.63	122.57	29.90	91.68
Adjusted R^2	0.017	0.055	0.018	0.050	0.016	0.039
Panel C3: Ordinary Least Squares						
$\log(1+\#\ \text{Visited Pages})_i$	0.016***	0.014***				
	(0.001)	(0.001)				
$\log(1+\#\ \text{App Launches})_i$			0.020***	0.017***		
			(0.001)	(0.001)		
$\log(1+\#\ \text{App Uses})_i$					0.027***	0.024***
					(0.001)	(0.001)
Adjusted R^2	0.016	0.020	0.016	0.020	0.017	0.022
Gender, Education, Job FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by City	Y	Y	Y	Y	Y	Y
Observations	98,293	49,271	98,293	49,271	98,293	49,271

Data Privacy and Digital Demand

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

Online Appendix

The Survey Questionnaire

Q1. Are you concerned about privacy issues while using online services?
Q2. What do you think about privacy protection in Alipay?
Q3. Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?
Q4. Will you avoid visiting mini-programs in Alipay because of privacy concerns?
Q5. What privacy issues are you concerned about when using mini-programs in Alipay? (You may select multiple choices.)

A. Data leakage and security;

B. Price discrimination by merchants;

C. Seductive advertising and temptation consumption;

D. Others

Q6. How many times will you agree if making authorization decisions for ten mini-programs?
Q7. How often do you regret authorizing information to mini-programs in Alipay?
Q8. Do you agree with the arguments below?

1) I agree to authorize data sharing with mini-programs since it is safe in Alipay.
2) I agree to authorize data sharing with mini-programs since my information has already been shared in many platforms.
3) I have to share my personal data in exchange for digital services even though I am concerned about my data privacy.
4) I authorize data sharing with a mini-program only when the requested information is not important.
5) I tend to authorize data sharing with mini-programs that are used by my friends.

Q9. Do you know how to change privacy settings in Alipay?
Q10. Have you ever changed your privacy settings in Alipay?
Q11. Do you know how to opt out from mini-programs in Alipay?
Q12. Have you ever opted out from mini-programs in Alipay?

Figure B1: Examples of the Authorization Page

This figure presents three examples of the authorization page with different information requirements. Users need to agree to share requested information before using mini-programs.

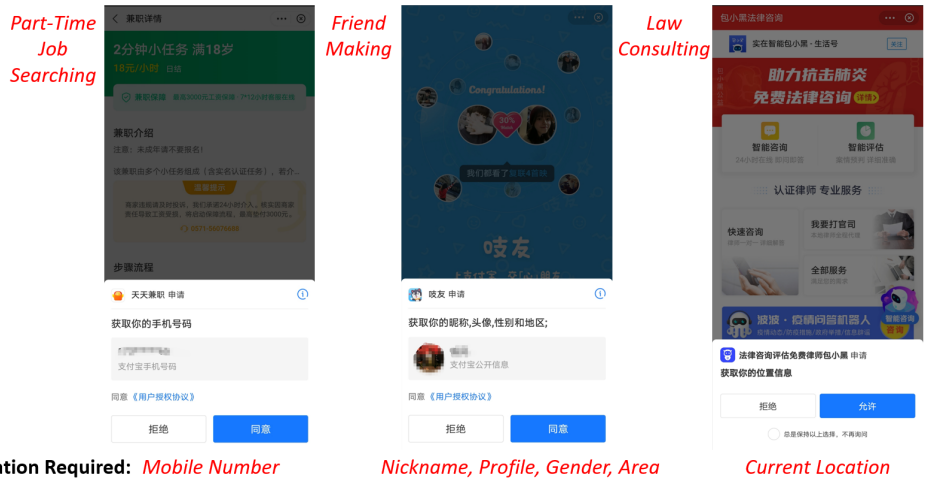


Figure B2: Completion Time Distribution (Seconds)

This figure plots the completion time distribution (seconds) in the survey conducted in July 2020 by Alipay. The vertical axis refers to the percentage of responses.

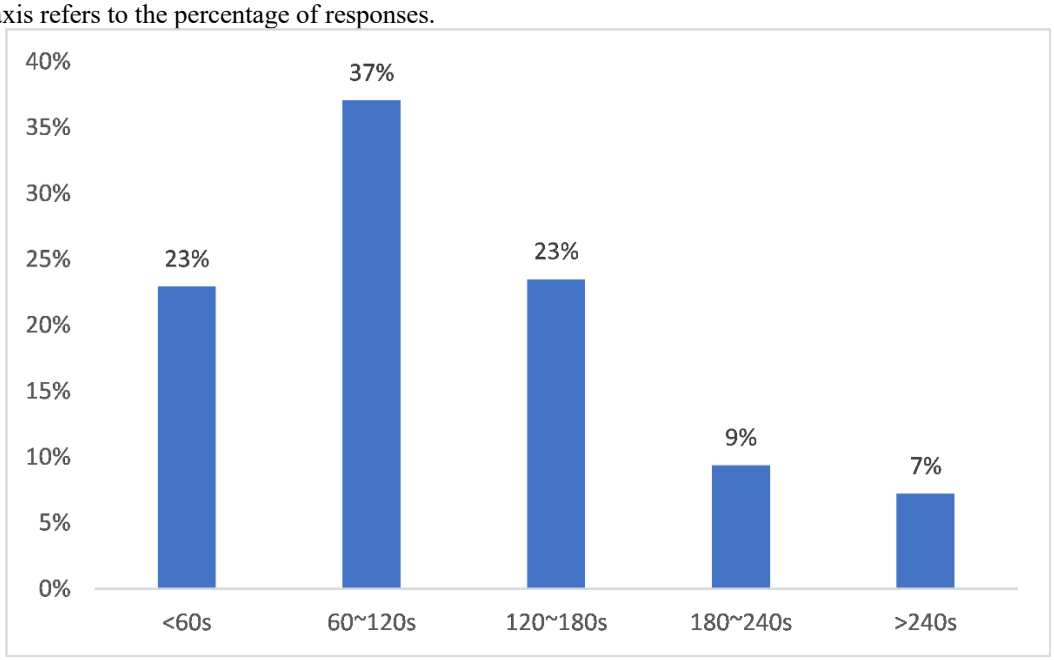


Figure B3: Age Distribution (Years)

This figure plots the age distribution of respondents from the survey conducted in July 2020 by Alipay. The vertical axis refers to the percentage of responses.

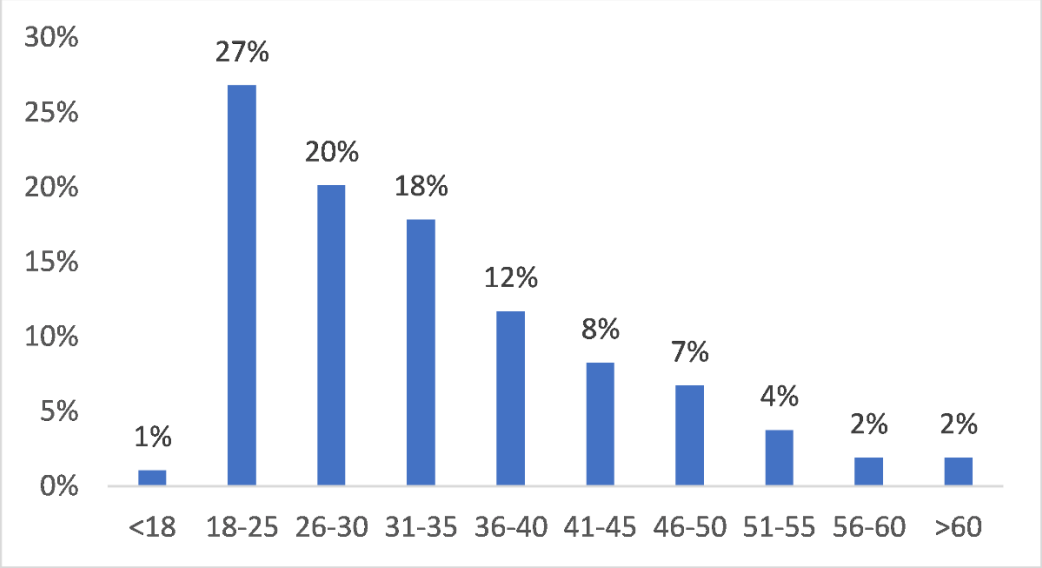


Figure B4: Distribution Across Provinces

This figure plots the linear correlation between distribution of respondents in the survey and distribution of the population across provinces. The vertical axis refers to the percentage of responses; the horizonal axis refers to the percentage of resident population (2019).

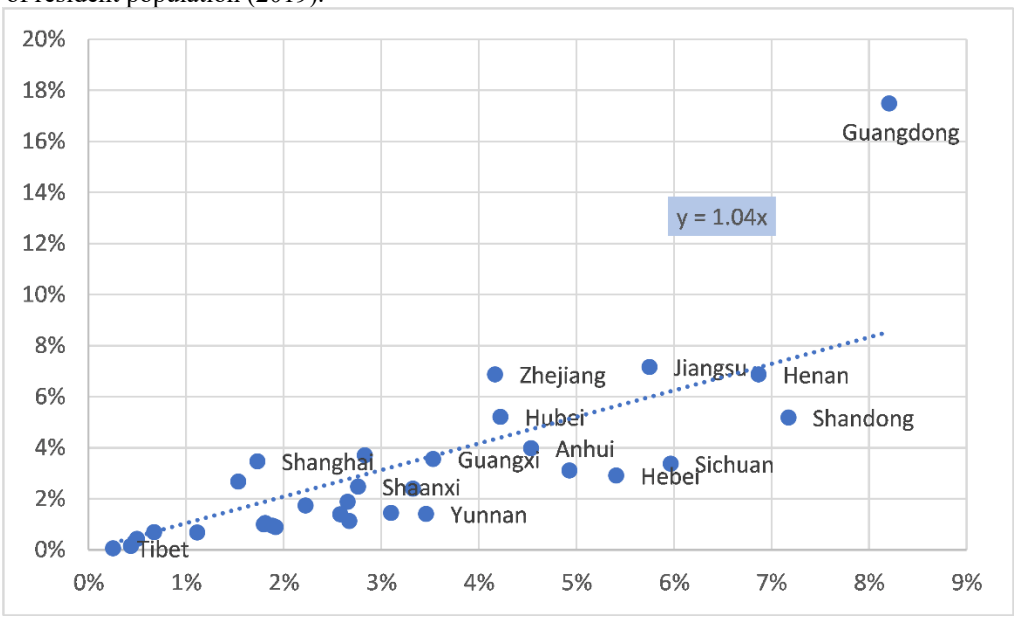


Figure B5: Digital Age Distribution (Years)

This figure plots the digital age distribution of respondents from the survey conducted in July 2020 by Alipay. The vertical axis refers to the percentage of responses; digital age is defined as the length of time since a user registered with Alipay.

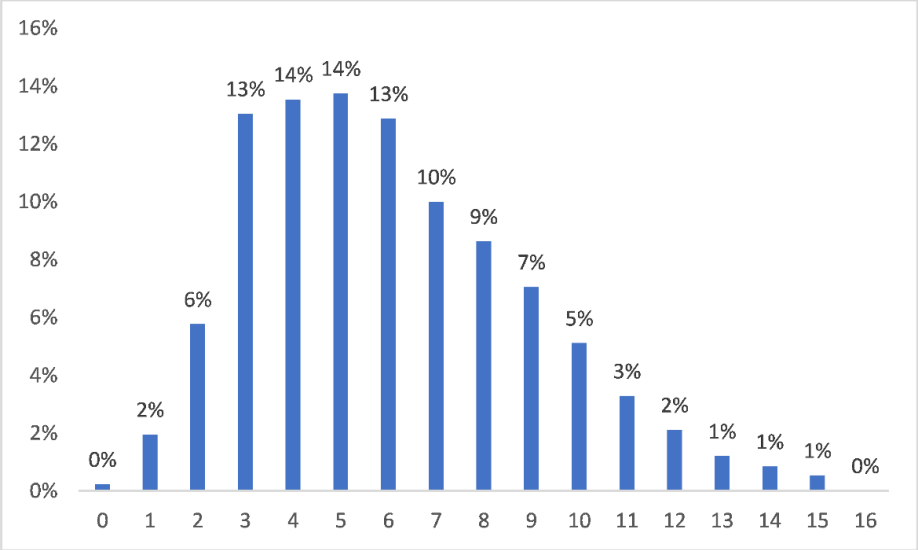


Figure B6: Cancellations During the 2017 Footprint Report Incident

The figure plots the fraction of Alipay users that canceled at least one mini-program authorizations in each day around the 2017 Annual User Footprint Report Incident, which happened on January 3, 2018.

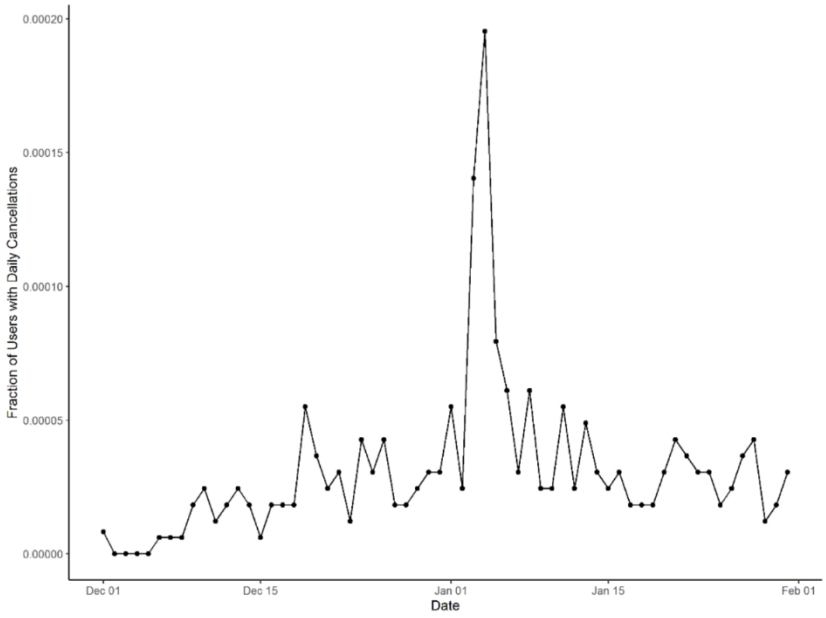


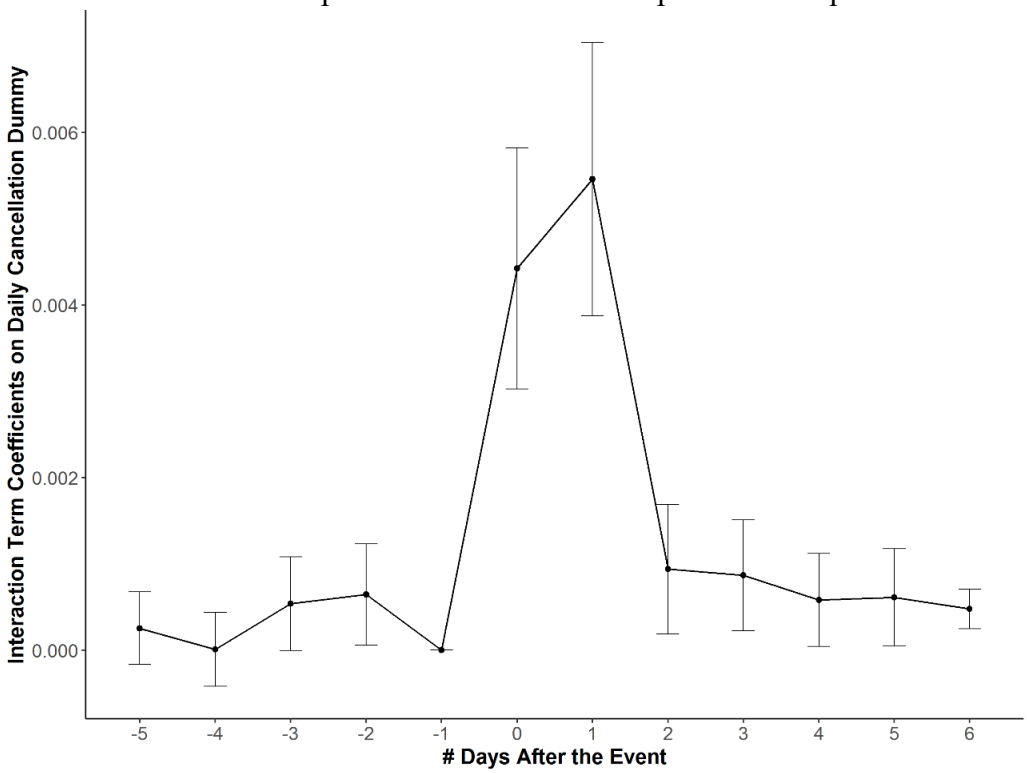
Figure B7: Activeness and Response to the 2017 Footprint Report Incident

The figures plot the $\beta_{H-L,\tau}$ coefficients estimated in the following regression with the bands indicating 95% confidence intervals:

$$\begin{aligned} &Daily\ Cancellation\ Dummy_{i,t} \\ &= \alpha_0 + \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{H-L,\tau} \cdot Heavy\ User_i \cdot \mathbb{I}(t=\tau) + \beta_{H-L,6} \cdot Heavy\ User_i \cdot \mathbb{I}(t \geq 6) \\ &+ \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_\tau \cdot \mathbb{I}(t=\tau) + \beta_6 \cdot \mathbb{I}(t \geq 6) + \delta_i + \varepsilon_{i,t}\ , \end{aligned}$$

where $Daily\ Cancellation\ Dummy_{i,t}$ is a dummy variable indicating whether user i has canceled at least one mini-program during day t; t corresponds to the number of days after January 3, 2018, (the day of the incident); $Heavy\ User_i$ is a dummy indicating whether the user i has used more mini-programs than 75% of the population as of November 30, 2017; δ_i is the individual fixed effects; and $\varepsilon_{i,t}$ is the error term that varies across individuals and over time. Panel A covers a representative random sample of 100,000 Alipay users without any filtering, and Panel B covers only the Alipay users who canceled at least one mini-program before November 30, 2017, in the representative random sample. The data are at individual and daily levels. The sample period ranges from December 29, 2017, through January 31, 2018.

Panel A: Representative Random Sample: Full Sample



Panel B: Representative Random Sample: Users Having Canceled Before November 30, 2017

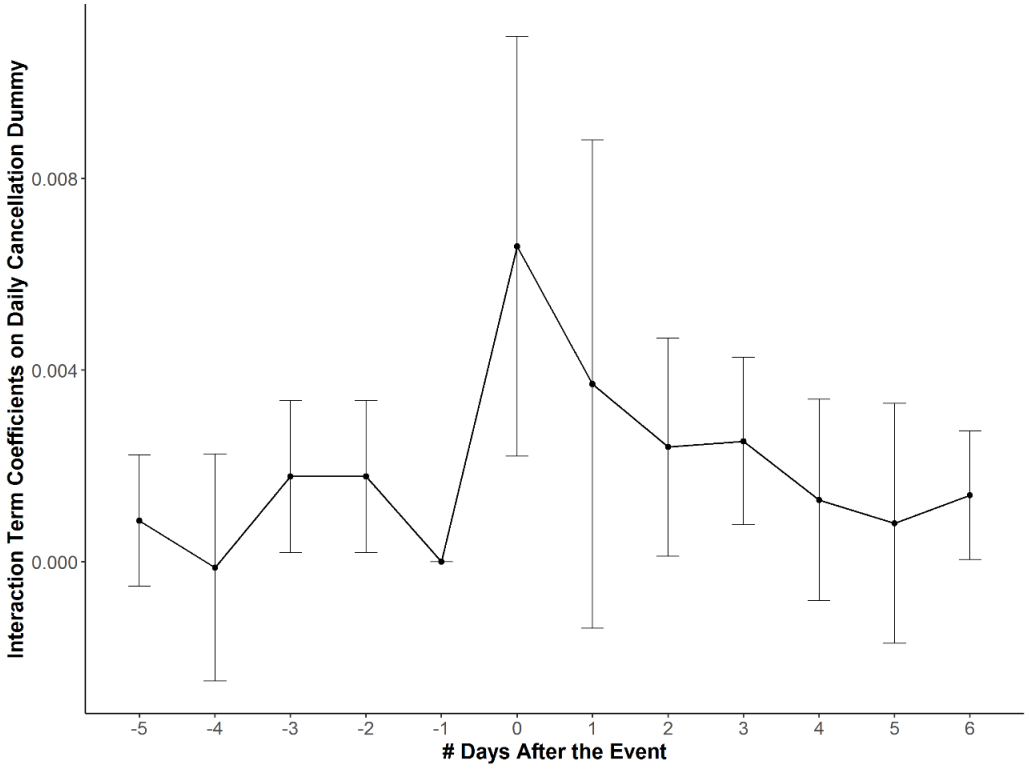
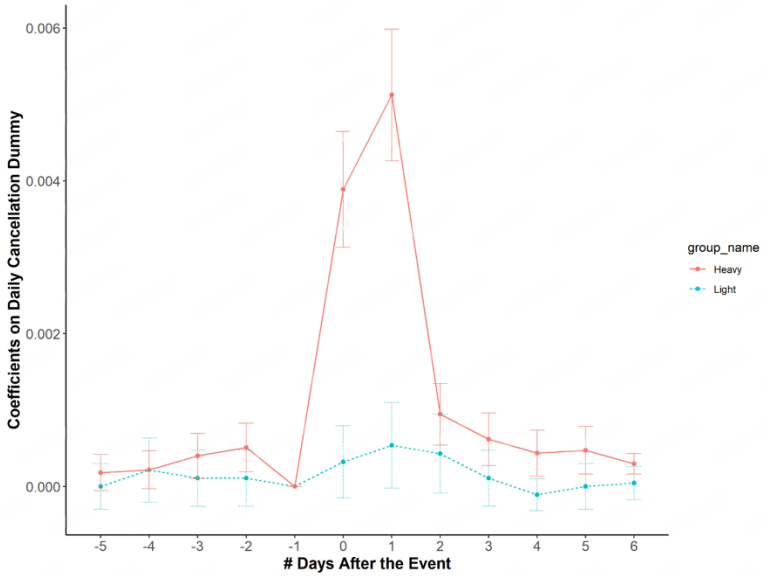


Figure B8: Alternative Activeness Measure and Response to the 2017 Incident

The figures plot the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated by the regression specified in Equation (4), where the bands indicate 95% confidence intervals. $Heavy\ User_i$ is a dummy indicating whether the user i has used more mini-programs than 50% of the population as of November 30, 2017. Panel A covers the random sample of 100,000 Alipay users without any filtering, and Panel B covers only the users who had canceled data sharing with at least one mini-program before November 30, 2017, in the random sample. The data are at individual and daily levels. The sample period ranges from December 29, 2017 to January 31, 2018.

Panel A: Unfiltered Users



Panel B: Users with Cancellation before November 30, 2017

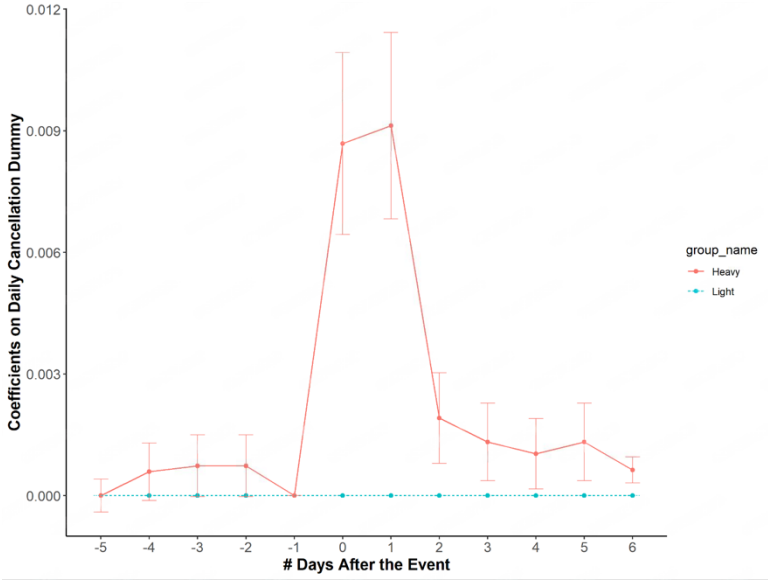


Table B1: Determinants of Data-Sharing Authorizations in Survey

Panel A summarizes the responses of the respondents to five statements. The respondents are split into two groups, one for those whose answers to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" were "concerned" or "very concerned," and the other group for those whose answers to this survey question were "not concerned." Panel B shows the regression results. The dependent variable takes a value of 1 if a respondent agrees with a statement. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: Summary of Responses to Survey Statements

	Count	Share	Count	Share	Total
	Agree		Disagree		
Q1: I agree to authorize data sharing with mini-programs because it is safe in Alipay.					
Concerned or very concerned	3,918	42%	5,331	58%	9,249
Not concerned	1,308	80%	318	20%	1,626
Q2: I agree to authorize data sharing with mini-programs because my information has already been shared in many platforms.					
Concerned or very concerned	1,083	12%	8,166	88%	9,249
Not concerned	493	30%	1,133	70%	1,626
Q3: I have to share my information in exchange for digital services even though I have concerns about my data privacy.					
Concerned or very concerned	6,030	65%	3,219	35%	9,249
Not concerned	913	56%	713	44%	1,626
Q4: I only authorize data sharing with mini-programs when the requested data are not important.					
Concerned or very concerned	1,852	20%	7,397	80%	9,249
Not concerned	485	30%	1,141	70%	1,626
Q5: I tend to authorize data sharing with mini-programs that are used by my friends.					
Concerned or very concerned	4,042	44%	5,207	56%	9,249
Not concerned	942	58%	684	42%	1,626

Panel B: Regression Analysis

Agree with	Q1	Q2	Q3	Q4	Q5
	(1)	(2)	(3)	(4)	(5)
Concerned or Very Concerned$_i$	-0.320***	-0.203***	0.083***	-0.096***	-0.158***
	(0.011)	(0.013)	(0.014)	(0.014)	(0.014)
Digital Experience$_i$	-0.001***	-0.001***	0.0003**	-0.001***	-0.00001
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Age$_i$	0.002***	0.001**	0.0005	0.004***	-0.001
	(0.001)	(0.0004)	(0.0005)	(0.0005)	(0.001)
City, Gender FE	Y	Y	Y	Y	Y
Observations	8,658	9,637	9,780	9,356	9,110
Adjusted R^2	0.070	0.052	0.013	0.019	0.014

Table B2: Responses to Other Survey Questions

This table summarizes responses to five survey questions regarding privacy concerns, data sharing, and opting out of mini-programs.

	Count	Share
Will you avoid visiting mini-programs in Alipay because of privacy concerns?		
Never	1,232	11%
Sometimes	5,876	54%
Often	3,767	35%
How many times will you agree if making authorization decisions for ten mini-programs?		
0-2	4,677	43%
3-4	3,210	30%
5-6	1,551	14%
7-8	710	7%
9-10	727	7%
How often do you regret authorizing information to mini-programs in Alipay?		
Never	3,214	30%
Sometimes	6,966	64%
Often	695	6%
Do you know how to opt out from mini-programs in Alipay?		
No	5,799	53%
Yes	5,076	47%
Have you ever opted out from mini-programs in Alipay?		
Maybe	2,953	27%
No	3,681	34%
Yes	4,241	39%

Table B3: Validating Selected Survey Responses

This table reports OLS regression and correlation results showing how survey responses relate to privacy-seeking actions, including data-sharing authorizations with mini-programs, changes to Alipay's default privacy settings, and authorization cancellations. *Authorization Rates in Survey* represents the response to the survey question: "*How many times will you agree if making authorization decisions for ten mini-programs?*". *Privacy Setting Changed in Survey* is a dummy variable equal to 1 if the respondent answered "yes" to the survey question "*Have you ever changed your privacy settings in Alipay?*" *Has Canceled in Survey* is a dummy variable equal to 1 if the respondent answered "yes" to the survey question "*Have you ever opted out from mini-programs in Alipay?*" *# Authorized Mini-Programs* is the total number of authorized mini-programs over both the pre-survey period (July 2019–July 2020) and the post-survey period (August 2020–December 2021). *Privacy Setting Changed* is a dummy variable indicating whether a user modified Alipay's default privacy settings between May 2017 and April 2020. *Has Canceled* is a dummy variable indicating whether a user revoked at least one data-sharing authorization between January 2013 and July 2020. Statistical significance is denoted as ***, **, and * for the 1%, 5%, and 10% confidence levels, respectively. Standard errors are reported in parentheses.

	# Authorized Mini-Programs$_i$	*Privacy Setting Changed*$_i$	*Has Canceled*$_i$
	(1)	(2)	(3)
Authorization Rates in Survey$_i$	0.147***		
	(0.009)		
Privacy Setting Changed in Survey$_i$		0.188***	
		(0.008)	
Has Canceled in Survey$_i$			0.197***
			(0.010)
Constant	0.712***	0.418***	0.404***
	(0.004)	(0.005)	(0.006)
Correlation between variables	0.165***	0.183***	0.192***
Observations	10,794	14,250	10,857
R^2	0.027	0.034	0.037

Table B4: The Data Privacy Paradox with Concern Types

These tables present regression analysis of the data privacy paradox results for the pre-survey period from July 2019 through July 2020, where the specific concern types are included in the regressions. *Concerned Dummy$_i$* and *Very Concerned Dummy$_i$* are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." we We include more types of privacy risks from the survey as controls—namely data leakage and security (*Concern Type 1$_i$*), price discrimination (*Concern Type 2$_i$*), and seductive advertising (*Concern Type 3$_i$*). *Education$_i$* is a dummy indicating whether the user has a college degree or higher. *Self Control$_i$* is a dummy indicating whether the user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the pre-survey period. Panel A displays results for user-month-level regressions. Columns (1)–(2) present results for the monthly number of authorized mini-programs in the full sample, columns (3)–(4) for the monthly number of initially visited mini-programs in the full sample, and columns (5)–(6) for the monthly number of authorized mini-programs in the subsample with nonzero monthly initial visits. *# Authorized Mini-Programs$_{it}$* and *# Visited Mini-Programs$_{it}$* are the number of mini-programs authorized and visited by individual *i* at time *t*, respectively. Panel B provides results for regressions at the user-mini–program level. For each user–mini-program pair, columns (1)–(2) report results for the dummy indicating whether the user allowed authorization, columns (3)–(4) for the dummy indicating whether the user visited at least once, and columns (5)–(6) for the dummy indicating whether the user allowed authorization, conditional on pairs where the user visited at least once. *Authorized Dummy$_{ij}$* equals one if a user *i* authorized data sharing to a mini-program *j*, otherwise zero. *Visited Dummy$_{ij}$* equals one if a user *i* visited a mini-programs at least once *j*, otherwise zero. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User-Month-Level Analysis

	# Authorized Mini-Programs$_{it}$		# Visited Mini-Programs$_{it}$		# Authorized Mini-Programs$_{it}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$	0.015	0.012	0.083***	0.071***	-0.042	-0.029
	(0.016)	(0.016)	(0.024)	(0.023)	(0.028)	(0.027)
Very Concerned Dummy$_i$	-0.001	-0.003	0.135***	0.118***	-0.098***	-0.079***
	(0.016)	(0.016)	(0.025)	(0.024)	(0.029)	(0.028)
Concern Type 1$_i$	-0.003	-0.014	0.011	0.041*	-0.030	-0.045
	(0.017)	(0.017)	(0.025)	(0.025)	(0.030)	(0.030)
Concern Type 2$_i$	0.064***	0.045***	0.005	0.054**	0.096***	0.046*
	(0.014)	(0.014)	(0.022)	(0.022)	(0.024)	(0.024)
Concern Type 3$_i$	0.033***	0.031***	0.044**	0.064***	0.055***	0.054***
	(0.012)	(0.012)	(0.019)	(0.019)	(0.021)	(0.021)
Digital Experience$_{it}$		0.001***		-0.0003		0.0002
		(0.0002)		(0.0003)		(0.0003)
Age$_{it}$		-0.002***		0.015***		-0.011***
		(0.001)		(0.001)		(0.001)
Constant	0.780***		0.997***		1.781***	
	(0.018)		(0.028)		(0.031)	
City, Gender, Time FE	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	152,110	151,886	152,110	151,886	70,988	70,892
Adjusted R^2	0.001	0.006	0.001	0.013	0.001	0.017

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$		Visited Dummy$_{ij}$		Authorized Dummy$_{ij}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Concerned Dummy$_i$ ($\times 10^{-4}$)	0.480	0.333	2.521***	2.137**	-321.864***	-260.269***
	(0.772)	(0.762)	(0.879)	(0.864)	(58.473)	(48.816)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	-0.474	-0.523	3.279***	2.788***	-617.986***	-492.727***
	(0.785)	(0.774)	(0.902)	(0.887)	(59.653)	(49.861)
Concern Type 1$_i$ ($\times 10^{-4}$)	-0.191	-0.856	0.255	0.554	-78.327	-176.726***
	(0.803)	(0.822)	(0.905)	(0.926)	(61.125)	(51.108)
Concern Type 2$_i$ ($\times 10^{-4}$)	3.830***	2.550***	2.469***	2.811***	339.424***	50.106
	(0.693)	(0.692)	(0.792)	(0.788)	(50.686)	(43.259)
Concern Type 3$_i$ ($\times 10^{-4}$)	1.214**	1.104*	1.353**	1.703***	23.935	20.158
	(0.571)	(0.571)	(0.658)	(0.655)	(45.392)	(37.704)
Digital Experience$_i$		5.130***		3.187***		275.220***
		(0.817)		(0.982)		(54.027)
Age$_i$		-1.757***		2.674***		-522.943***
		(0.294)		(0.375)		(19.296)
Constant	0.004***		0.005***		0.851***	
	(0.0001)		(0.0001)		(0.006)	
Mini-program, City and Gender FE	N	Y	N	Y	N	Y
Cluster by Individual	Y	Y	Y	Y	Y	Y
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Visited Only	Visited Only
Observations	25,414,875	25,364,288	25,414,875	25,364,288	132,924	132,713
Adjusted R^2	0.000	0.105	0.000	0.129	0.004	0.148

Table B5: Robustness: Causal Effect of Digital Demand on Privacy Concerns

This table examines the causal relationship between digital demand and revealed privacy concern in the survey, utilizing an alternative instrumental variable that accounts for city size as a robustness check. *Prior Bike Placement Per Capita*$_c$ is the average monthly number of Alipay-bundled shared bicycles placed in the user *i*'s city *c*, normalized by the number of active Alipay users before the survey in July 2020. The dependent variable, *Concerned Dummy*$_i$, is a dummy variable that equals 1 if the user *i*'s answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." We use the number of visited pages in columns (1)–(2), the number of mini-programs launches in columns (3)–(4), and the number of mini-programs uses in columns (5)–(6) to capture the demand for digital services. All usage variables are calculated by summing the activities for each individual across all mini-programs during the pre-survey period from July 2019 through July 2020. Panel 1 outlines the 2SLS estimates; Panel 2 describes the first stage; Panel 3 provides the OLS results. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Concerned Dummy$_i$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: Two-Stage Least Squares						
$\log(1 + \#\ \widehat{\text{Visited Pages}})_i$	0.141***	0.085**				
	(0.035)	(0.036)				
$\log(1 + \#\ \widehat{\text{App Launches}})_i$			0.171***	0.107***		
			(0.043)	(0.046)		
$\log(1 + \widehat{\#\ \text{App Uses}})_i$					0.180***	0.118***
					(0.049)	(0.054)
Panel 2: First Stage for Usage Variables						
Prior Bike Placement Per Capita$_c$	7.914***	8.341***	6.523***	6.637***	6.219***	6.013***
	(1.530)	(1.600)	(1.411)	(1.452)	(1.522)	(1.481)
F-Statistic	26.75	17.00	21.36	25.95	16.70	42.55
Adjusted R^2	0.006	0.024	0.005	0.031	0.005	0.050
Panel 3: Ordinary Least Squares						
$\log(1+\#\ \text{Visited Pages})_i$	0.013***	0.014***				
	(0.002)	(0.003)				
$\log(1+\#\ \text{App Launches})_i$			0.014***	0.015***		
			(0.002)	(0.003)		
$\log(1+\#\ \text{App Uses})_i$					0.015***	0.016***
					(0.002)	(0.003)
Adjusted R^2	0.003	0.034	0.003	0.034	0.004	0.034
Gender, Education, Job FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster by City	Y	Y	Y	Y	Y	Y
Observations	10,871	6,785	10,871	6,785	10,871	6,785

Table B6: Summary Statistics of the Panel Data with Credit Line Information

This table reports summary statistics of the user-month panel data in Ouyang (2022) of 41,485 Alipay users from May 2017 to September 2020. The variables are categorized into three types at different levels. At the individual level, *Is Male*$_i$ is a dummy for male; *Low Education*$_i$ is a dummy for below bachelor's degree; *Bike User*$_i$ is a dummy for using a shared bike at least once. At the city-month level, *log(Bike Placement)*$_{c,t}$ is the log number of active shared bikes in city cat time t. At the individual-month level, *log(1+Credit Line)*$_{i,t}$ is the *log(1+x)* transformed credit line; *Credit Access*$_{i,t}$ is a dummy for having access to Alipay's virtual credit card at time t; *log(Credit Line)*$_{i,t}$ is the log credit line conditional on *Credit Access*$_{i,t}$; *log(In-Person Payment Flow)*$_{i,t}$ is the log amount of in-person payments using Alipay.

	N	Mean	Std	Min	p25	Median	p75	Max
Individual Level								
Is Male$_i$	41,214	0.54	0.50	0.00	0.00	1.00	1.00	1.00
Low Education$_i$	41,459	0.88	0.33	0.00	1.00	1.00	1.00	1.00
Bike User$_i$	41,485	0.29	0.45	0.00	0.00	0.00	1.00	1.00
City-Month Level								
log(Bike Placement)$_{c,t}$	12,665	7.08	3.39	0.00	4.11	7.85	9.91	13.91
Individual-Month Level								
log(1+Credit Line)$_{i,t}$	1,321,837	4.89	4.02	0.00	0.00	6.55	8.52	11.02
Credit Access$_{i,t}$	1,321,837	0.62	0.49	0.00	0.00	1.00	1.00	1.00
log(Credit Line)$_{i,t}$	819,812	7.88	1.58	3.00	6.91	8.13	9.13	11.02
log(In-Person Payment Flow)$_{i,t}$	688,428	5.70	2.29	-4.61	4.31	6.04	7.27	15.88

Table B7: Digital Footprints, Data Sharing & Nonlinear Credit Effects

This table examines the nonlinear relationship between mini-programs data sharing and financial credit access in Alipay over multiple future periods. We categorize both the number of authorized and canceled mini-programs into mutually exclusive bins, allowing for flexible, nonparametric modeling of nonlinear effects. The table reports Poisson regression results with the dependent variable being the credit line granted at periods t, t+1, ..., up to t+6. Key independent variables include indicator dummies for bins of the Number of Authorized Mini-Programs (1, 2, 3-5, 6-10, and 10+) and Canceled Mini-Programs (1, 2, 3-5, 6-10, and 10+) at the individual-month level. All specifications control for digital payment amount using absorbed dummies for each centile of this variable. The results for regressions are at the user-month level. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. Standard errors, clustered at the individual and time levels, are reported in parentheses.

	Credit Line$_{iT}$						
	t	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Authorized Mini-Programs$_{it}$: 1	0.650***	0.626***	0.596***	0.562***	0.530***	0.499***	0.468***
	(0.051)	(0.048)	(0.046)	(0.044)	(0.042)	(0.041)	(0.041)
# Authorized Mini-Programs$_{it}$: 2	0.990***	0.953***	0.908***	0.859***	0.809***	0.761***	0.713***
	(0.076)	(0.071)	(0.068)	(0.065)	(0.063)	(0.061)	(0.060)
# Authorized Mini-Programs$_{it}$: 3-5	1.298***	1.240***	1.172***	1.100***	1.032***	0.967***	0.902***
	(0.090)	(0.084)	(0.081)	(0.078)	(0.075)	(0.073)	(0.072)
# Authorized Mini-Programs$_{it}$: 6-10	1.537***	1.458***	1.370***	1.278***	1.191***	1.109***	1.027***
	(0.101)	(0.094)	(0.090)	(0.087)	(0.084)	(0.082)	(0.080)
# Authorized Mini-Programs$_{it}$: 10+	1.612***	1.518***	1.415***	1.309***	1.209***	1.116***	1.025***
	(0.103)	(0.096)	(0.092)	(0.087)	(0.084)	(0.082)	(0.080)
# Canceled Mini-Programs$_{it}$: 1	-0.002	-0.002	-0.002	-0.001	0.001	0.001	0.001
	(0.029)	(0.029)	(0.028)	(0.028)	(0.027)	(0.027)	(0.026)
# Canceled Mini-Programs$_{it}$: 2	-0.063*	-0.065*	-0.064**	-0.063**	-0.061*	-0.058*	-0.056*
	(0.034)	(0.033)	(0.032)	(0.032)	(0.032)	(0.032)	(0.031)
# Canceled Mini-Programs$_{it}$: 3-5	-0.139***	-0.139***	-0.140***	-0.141***	-0.140***	-0.140***	-0.137***
	(0.035)	(0.035)	(0.034)	(0.033)	(0.032)	(0.032)	(0.031)
# Canceled Mini-Programs$_{it}$: 6-10	-0.198***	-0.192***	-0.185***	-0.178***	-0.172***	-0.169***	-0.166***
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
# Canceled Mini-Programs$_{it}$: 10+	-0.351***	-0.363***	-0.372***	-0.373***	-0.372***	-0.372***	-0.372***
	(0.081)	(0.081)	(0.080)	(0.080)	(0.081)	(0.082)	(0.083)
Individual, Time FE	Y	Y	Y	Y	Y	Y	Y
Digital Payment Amount Centiles Dummies	Y	Y	Y	Y	Y	Y	Y
Cluster by Individual, Time	Y	Y	Y	Y	Y	Y	Y
Observations	1,015,868	1,002,584	972,894	943,285	913,751	884,296	854,952
Adjusted Pseudo R^2	0.878	0.877	0.879	0.881	0.883	0.886	0.888

Table B8: Summary Statistics of the Random Sample

This table reports summary statistics of a representative random sample of 100,000 Alipay users. Panel A reports user information in three parts. The first part reports the general information. *Privacy Setting Changed*, a proxy measure for privacy concerns, is a dummy variable equal to 1 if a user changed their privacy setting at least once between May 2017, and April 2020, and 0 otherwise. *Digital Experience* is the number of months since the user first registered on Alipay, and *Age* is the user's physical age in July 2020. The second part covers data sharing with mini programs, including the number of authorized and visited mini-programs over the pre-survey period of July 2019 to July 2020; the *Has Canceled*$_i$ status and # *Cancellations*$_i$, of used mini-programs over the pre-survey period of January 2013 to July 2020. The third part reports summary statistics of monthly use variables of Alipay users in each mini-program during the pre-survey period from July 2019 to July 2020, including number of uses, number of launches, and number of visited pages. Use variables are winsorized at the 1% and 99% levels.

	N	Mean	Std	Min	p25	Median	p75	Max
General Information								
Privacy Setting Changed$_i$	98,679	0.09	0.28	0.00	0.00	0.00	0.00	1.00
Digital Experience$_i$ (month)	99,600	60.69	36.81	0.00	32.00	55.00	82.00	190.00
Age$_i$ (year)	97,876	36.61	12.89	1.00	27.00	34.00	46.00	120.00
Data Sharing with Mini-Programs								
# Authorized Mini-Programs$_i$	100,000	2.40	3.52	0.00	0.00	1.00	3.00	136.00
# Visited Mini-Programs$_i$	100,000	3.02	4.59	0.00	0.00	2.00	4.00	248.00
Has Canceled$_i$	99,995	0.12	0.32	0.00	0.00	0.00	0.00	1.00
# Cancellations$_i$	98,674	0.30	1.45	0.00	0.00	0.00	0.00	61.00
Monthly Mini-program Use								
# App Uses$_{it}$	3,036,555	0.34	2.21	0.00	0.00	0.00	0.00	40.00
# App Launches$_{it}$	3,036,555	1.10	6.90	0.00	0.00	0.00	0.00	123.00
# Visited Pages$_{it}$	3,036,555	3.06	19.96	0.00	0.00	0.00	0.00	342.00