Data Privacy and Digital Demand*

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

February 2025

Abstract

We combine survey and behavioral data to analyze consumers' data-sharing choices within a major financial platform. Paradoxically, respondents with stronger privacy concerns authorize more data sharing, illustrating the data privacy paradox. We resolve this paradox as an omitted variable problem: users with stronger privacy concerns also exhibit higher demand for digital services. Moreover, our analysis uncovers a nuanced interplay between digital demand, data sharing, and privacy concerns—greater reliance on digital services heightens privacy concerns, yet the data shared during usage enables providers to enhance these services, reinforcing data sharing despite growing privacy concerns.

* 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 appreciate comments and suggestions made by Alessandro Acquisti, Laura Brandimarte, Avi Goldfarb, Zhuang Liu, Cameron Peng, Avanidhar Subrahmanyam, Huan Tang, Catherine Tucker, Liad Wagman, David Yang, Liyan Yang, and 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 grateful for the help provided by Yong Li and Dongyu Wei at the Luohan Academy and Xiaopeng Wang and Xiaoqiang Yang at Ant Group. 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., 2022; Ouyang, 2022), investment recommendations (Rossi and Utkus, 2020), and insurance underwriting (Fang et al., 2020)— highlights its far-reaching implications.[1]

More broadly, theoretical frameworks by Jones and Tonetti (2020), Farboodi and Veldkamp (2020), and Cong, Xie, and Zhang (2020) emphasize the central role of consumer data sharing in driving financial innovation and supporting the expanding data economy, positioning it as a cornerstone of the broader macroeconomy.[2] While this transformation unlocks new financial opportunities, it also raises pressing concerns about data privacy. The growing reliance of digital platforms and AI systems, such as Amazon, ChatGPT and Alipay, on user data has further amplified these concerns.[3]

Despite the critical role of data sharing, the intricacies of consumers' decisions in this domain remain largely unexplored. A key aspect of this discussion is the data privacy paradox, where consumers' data-sharing behaviors contradict their stated privacy concerns—a phenomenon documented by Spiekermann, Grossklags, and Berendt (2001), Gross and Acquisti (2005), Norberg, Horne, and Horne (2007), and Athey, Catalini, and Tucker (2017). This paradox underscores the need for a deeper understanding of how consumers navigate data-sharing decisions in real-world settings. Its persistence could expose consumers to unforeseen financial risks when engaging with digital platforms. Conversely, heightened privacy concerns may discourage data sharing, potentially limiting access to valuable financial services that rely on consumer data, such as inclusive lending or personalized wealth management solutions.

We analyze the data privacy paradox within a major finance platform using a unique dataset that links consumers' privacy attitudes with their data-sharing choices. First, we assess whether the paradox holds in real-world settings where consumers decide whether to share data with digital

[1] Berg, Fuster, and Puri (2021) review the expanding literature on FinTech lending, covering market size, key innovations in customer-lender interactions, and the use of alternative data and machine learning for screening. They highlight how repurposing consumer data for financial applications raises significant privacy concerns while also enhancing access to credit and other financial services.

[2] For a comprehensive review of data sharing in the growing data economy, see Chen et al. (2021).

[3] This shift in consumer attitudes is evident in 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.

service providers. Solove (2021) questions the reliability of self-reported privacy concerns, which often stem from broad, abstract questions rather than specific decisions consumers face. Next, we explore whether the gap between stated concerns and actual data-sharing behavior reflects a lack of informed decision-making. Finally, we examine the key factors driving consumers' privacy concerns and their choices to share or withhold personal data.

Specifically, we analyze users' data-sharing choices on Alipay, a finance platform operated by Ant Group. Originally launched in 2004 as an online payment service, Alipay has evolved into a comprehensive digital finance ecosystem, offering payments, credit, insurance, and investments. A key component is its mini-programs, introduced in 2017—lightweight applications within the Alipay app that provide various services without requiring separate downloads. With over 2 million mini-programs spanning finance, e-commerce, and local utilities, third-party providers leverage Alipay's vast user base and data infrastructure to deliver tailored services.

To access a mini-program, users must first consent to sharing certain personal data, including both initial information and data generated during use. The scope of data sharing varies widely, from basic details like an Alipay nickname to more sensitive information such as national ID numbers and transaction histories. The diversity of mini-programs—both in the value of services they provide and the sensitivity of the data they request—creates an ideal setting to examine how users navigate the trade-offs between privacy concerns and the benefits of digital services.

Our study surveyed 14,250 Alipay users in July 2020 to assess their concerns about sharing data with mini-programs and linked their responses to administrative data capturing both their 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 shared data at similar levels. This pattern holds even after controlling for user characteristics such as digital savviness, age, gender, city of residence, and mini-program attributes. Importantly, our administrative data confirms the reliability of self-reported privacy concerns, as users' propensity to cancel prior data-sharing authorizations monotonically increases with their stated privacy concerns. These findings

underscore the data privacy paradox, even in the absence of measurement error, within Alipay's data-driven financial ecosystem.

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 reversing the earlier finding of a privacy paradox.

This surprising role of the omitted variable underscores the need to fully understand the complex relationship between privacy concerns and digital demand. Two key mechanisms operate in opposing directions. First, we examine how intensified digital engagement leads to stronger privacy concerns. Greater reliance on digital applications enables service providers to accumulate vast amounts of user data, heightening risks of data breaches (Fainmesser et al., 2019), price discrimination (Taylor, 2004; Acquisti and Varian, 2005), and exploitation of personal vulnerabilities (Liu et al., 2020), all of which amplify users' privacy concerns.

To establish causality, 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, interacted with the user's prior experience with shared bikes. 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.

Given that privacy concerns intensify with digital engagement, one might expect users to become more reluctant to share data over time. However, we find the opposite—users authorized more data sharing post-survey, with the increase being particularly pronounced among those who

had previously expressed privacy concerns. While unexpected, this pattern—once again reflecting the data privacy paradox—suggests that growing digital demand, likely driven by more convenient services, outweighs heightened privacy concerns, ultimately leading to greater data sharing.

This finding highlights the other side of the relationship between digital demand and privacy concerns: data sharing driven by greater digital engagement enhances service providers' ability to improve digital applications, further fueling digital demand. As consumers share more data, service providers refine personalized services, making digital applications even more indispensable (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020; Cong et al., 2020).

To address potential bias from the likelihood that more active Alipay users responded to our survey, we analyzed a representative sample of 100,000 users, randomly selected from all active Alipay users. Using a behavior-based measure of privacy concerns—whether a user modified Alipay's default privacy settings—we confirmed the key findings from our survey sample. Specifically, users with stronger privacy concerns approved more data sharing with mini-programs, but this paradox diminishes after accounting for digital demand. Additionally, intensified digital engagement leads to stronger privacy concerns.

Leveraging this representative sample, we also examined how privacy concerns evolved after a notable incident on January 3, 2018, which heightened awareness of data privacy among Alipay users. Notably, this event made active mini-program users more prone to privacy concerns, further reinforcing the link between greater digital reliance and heightened privacy concerns.

Using a representative sample of Alipay users from Ouyang (2022) with credit access data, we also find that users who authorize more mini-programs gain greater credit access, while those who frequently cancel data sharing may forgo potential credit. These findings highlight the direct link between data sharing and credit access.

Our study not only validates and resolves the data privacy paradox within a major finance platform but also leverages this paradox to examine a key dynamic of the digital ecosystem: the simultaneous growth of consumers' privacy concerns and digital demands. This dual perspective enhances our understanding of both the costs and benefits of data sharing.

On the cost side, our analysis shows that privacy concerns intensify as consumers engage more with digital applications and share increasing amounts of personal data with service providers. This finding not only confirms the upward shift in privacy concerns (e.g., Goldfarb and Tucker, 2012) but also underscores a crucial characteristic of data privacy—it is not merely an isolated

preference, as sometimes framed in policy discussions, but a risk that arises through data sharing in digital interactions.

Economists have long emphasized that the value of privacy is tied to the economic consequences of concealing one's private type (Stigler, 1980; Posner, 1981). These consequences depend on the context in which consumer data is shared and with whom. While data sharing can enhance product matching by enabling sellers to better understand consumer preferences, it also exposes consumers to risks such as price discrimination (Taylor, 2004; Acquisti and Varian, 2005), data breaches or leaks (Fainmesser, Galeotti, and Momot, 2019), and targeted advertising that exploits consumer vulnerabilities (Liu, Sockin, and Xiong, 2020).

On the benefit side, the literature on the data economy (e.g., Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020; Cong, Xie, and Zhang, 2020) highlights two key features of data sharing—nonrivalry and increasing returns to scale. These characteristics suggest that once consumer data is shared, digital service providers can leverage it to enhance service quality and expand offerings, further driving consumer demand for digital services. The simultaneous rise in both the costs and benefits of data sharing is essential for explaining the increasing trend in Alipay users' data-sharing authorizations observed in our sample.[4]

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. 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 (2020) 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. (2022) 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.

[4] If privacy concerns outpace digital demand, they may constrain the growth of the data-sharing economy and data-driven financial services. Strengthening privacy protections is crucial to unlocking the full potential of data sharing. This concern has driven a growing literature on the impact of data privacy regulations (e.g., Goldberg, Johnson, and Shriver, 2019; Aridor, Che, and Salz, 2020) and inspired decentralized digital platforms using cryptographic technologies to curb platform dominance over consumer data (Sockin and Xiong, 2022).

The paper is structured as follows: Section I outlines the institutional background of Alipay users' data sharing with mini-programs. Section II describes the user survey and summary statistics. Section III analyzes the data privacy paradox, while Section IV examines the link between privacy concerns and digital demand. Section V presents robustness checks, Section VI explores the relationship between data sharing and credit access, and Section VII concludes. Additional analyses are provided in the 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 and fintech platform. It serves over 900 million users in China, covering more than 70% of the population. As a major fintech platform, Alipay provides a wide range of financial services, including digital payments, small loans, credit cards, insurance, and wealth management. This section outlines the Alipay platform and its data-sharing framework, focusing on the relationship between Alipay users and third-party mini-programs.

Alipay has expanded its ecosystem with mini-programs, essentially 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. The share of Alipay users engaging with mini-programs grew from 21% in Q4 2015 to 49% in Q2 2019 (Chen et al., 2021), highlighting the rise of embedded finance and the seamless integration of financial services into everyday activities.

Using mini-programs involves an explicit exchange of personal financial data for access to digital services. Upon first visit, a mini-program requests user authorization to share specific information necessary for its services. [5] The requested data varies—ranging from basic details like

[5] For example, Hellobike, a widely used bike-sharing mini-program, can be accessed either through its standalone app or within Alipay. When users first visit the Hellobike mini-program, they are prompted to authorize three types of data sharing. The first includes basic information such as their nickname, profile picture, gender, and location. The second involves their credit score, which helps assess trustworthiness and determine whether a deposit is required. The third covers identification and transaction data, including their real name, phone number, national ID number, and transaction history within the mini-program. Once users authorize data sharing, they can access Hellobike's services. Figure A1 in the Online Appendix provides three additional examples. One mini-program facilitates job searches and requests users to share their mobile numbers. Another focuses on social connections and requires users to share their nickname, profile picture, gender, and location. The third provides legal consulting services and requests access to users' location data.

an Alipay nickname to more sensitive information such as a national ID number and transaction history. Users must either accept or reject the request, with full access granted only upon authorization. [6] Once authorized, users can cancel data sharing at any time before the authorization period ends. This study examines both authorization and cancellation decisions among a sample of Alipay users.

In Alipay's main app, users navigate various data-sharing settings, such as displaying their real name to friends, making recent posts public, allowing connections without prior permission, and being searchable by phone number. These options enable users to tailor their privacy preferences, though the default settings prioritize maximum visibility and connectivity. Some users choose to modify these defaults, signaling privacy concerns about sharing personal information with other Alipay users. In our analysis, we use changes to default privacy settings as a behavioral indicator of privacy inclination, complementing our survey-based measure of privacy concerns.

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

[6] Our setting presents a more straightforward trade-off compared to the data-sharing decisions consumers face on many public websites. Under the GDPR mandate, public websites allow users to opt in or out of data collection. Typically, if a user consents, the website can utilize their data to offer personalized services. However, even if the user opts out, they can still access the website, albeit without personalization. In this case, the trade-off involves choosing between personalized and non-personalized services. In contrast, Alipay users must authorize data sharing to access any services within a mini-program. Without authorization, they are unable to use the mini-program at all, making the decision a binary choice rather than a spectrum of service personalization.

with mini-programs. Among the respondents, 10,875 reported having used mini-programs, representing 76% of the total. [7] 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 Alipay users rather than the entire user base. For robustness and comparison, Section VI 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 compare data-sharing authorizations among respondents with different levels of concern about sharing data with mini-programs.

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 (2019). The second represents the risk that extensive data sharing enables merchants to infer consumers' reservation prices and engage in price discrimination—a concern widely

[7] Figures A2–A5 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 A2). The geographical distribution of the respondents across the provinces in China lines up well with the distribution of the population (see Figure A4), 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%.

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 worries of users.

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.

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.

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.

To capture privacy concerns, we construct dummy variables based on responses to the survey question: "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" Respondents could choose from "*not concerned*," "*concerned*," or "*very concerned*." The *Concerned Dummy* is assigned a value of 1 if the 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."

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 counts the number of initial visits to mini-programs, which trigger data-sharing requests. The second records how many times users

authorize these requests. The remaining three 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 dummy*, which takes a value of 1 if a user has 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* counts the number of active mini-programs from which a user has revoked data-sharing authorization during this period, where a mini-program is considered active if the user has engaged with it at least once. Finally, the *Cancellation Rate* is calculated as the number of canceled authorizations between January 2013 and July 2020, divided by the total number of active mini-programs.

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 have 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 revoke 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 four key variables: the number of active days, 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 day, launch it several times per session, and visit multiple pages within each launch. On average, each month, a respondent in our survey sample was active in a mini-program on 0.57 days, engaged 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*. While age shows no significant variation across groups, 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.

III. The Data Privacy Paradox

In this section, we examine how the respondents' data-sharing choices are related to their privacy concerns by combining their survey responses with administrative data. When deciding whether to share personal data with a specific mini-program, Alipay users weigh the benefits they receive against potential privacy risks. Both benefits and costs vary depending on the user and the particular mini-program in question.

To formalize this trade-off, we propose a linear decomposition of the cost for user i when sharing data with mini-program j, denoted as c_{ij}:

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

where c_i represents the user's inherent privacy concerns, c_j reflects the mini-program's privacy risk—particularly if it requires more sensitive data or has a questionable privacy protection reputation, and ϵ_{ij} is an idiosyncratic noise component independent of both the user and mini-program.

Similarly, we decompose the benefit b_{ij} that user i derives from mini-program j as:

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

where b_i captures the user's general inclination or receptiveness to digital services, b_j represents the value or utility offered by the mini-program's services, and ε_{ij} is another noise component, independent of the user-mini-program pairing.

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

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

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

To begin, we consider a baseline scenario where b_i and c_i are unrelated—meaning that a user's valuation of digital services is independent of their privacy concerns. This assumption aligns with common policy debates on data privacy, where privacy apprehensions are often treated in isolation from consumer demand for digital services. Based on this perspective, we posit:

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 data privacy paradox. In this section, we begin by empirically testing this hypothesis. Our initial results reject the hypothesis and instead provide evidence consistent with the data privacy paradox. We then validate our survey-based measure of privacy concerns, demonstrating that the paradox remains robust to measurement errors. Finally, we explore an alternative explanation, considering the possibility of a positive correlation between b_i and c_i across users. In this scenario, a user's demand for digital services may counterbalance their privacy concerns, making their data-sharing decisions less sensitive to those concerns.

A. Privacy Concerns and Data Sharing

In Figure 1, we compare the number of data-sharing authorizations by Alipay users who expressed different levels of concern about data sharing in their responses to the survey question, "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" Panel A shows that during the pre-survey period of July 2019 to July 2020, "unconcerned" users on average initially visited 14.3 mini-programs and authorized data sharing with 11.2 of them, "concerned" users visited 15.5 mini-programs and authorized 11.5, and "very concerned" users visited 16.3 mini-programs and authorized 11.3. There is an interesting pattern that "concerned" and "very concerned" users tend to open more new mini-programs than "unconcerned" users and eventually authorize data sharing with almost the same number of mini-programs. This pattern contradicts Hypothesis 1 that privacy-concerned users are more reluctant to authorize data sharing.

As users also differ in other dimensions beyond privacy concerns, we adopt a panel regression at the user-month level to control for various user characteristics, as well as time-varying effects:

$$Y_{it} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \lambda_t + \epsilon_i\ , \quad (1)$$

where the dependent variable Y_{it} is a measure of certain behavior (either the number of data-sharing authorizations or initial visits to mini-programs) by user i in month t; the dummy variable $Concerned_i$ is defined to be 1 if user i answers "concerned" to the question about sharing data with mini-programs in the survey, and 0 otherwise; the dummy variable $Very\ Concerned_i$ is defined to be 1 if user i answers "very concerned" in the corresponding question, and 0 otherwise;

Age_i and $Digital\ Experience_i$ are two control variables; δ_i represents fixed effects related to other user characteristics, including gender and city; and λ_t represents time fixed effects. Without including the controls, the sample size is 152,110. As the characteristics of some users are missing, including the control variables slightly reduces the sample size to 151,886.

Panel A in Table 3 reports the regression results, using the sample from July 2019 to July 2020. Columns (1) and (2) show that the estimates of a_1 and a_2 are both insignificant, with or without the controls, confirming that "concerned" and "very concerned" users do not authorize data sharing with fewer mini-programs than "unconcerned" users. Furthermore, columns (3) and (4) show that the level of privacy concerns is positively correlated with the number of initially visited mini-programs, even though it is uncorrelated with the number of data-sharing authorizations. Specifically, privacy-concerned users, on average, initially visit 0.088 more mini-programs per month, and "very concerned" users, on average, have 0.140 more initial visits per month; the coefficients are both highly significant.

A user's data-sharing authorization with a mini-program may also depend on the services offered and the data requested by the mini-program. To control for these mini-program characteristics, we further expand our regression analysis to the user-mini–program level for all possible pairs of users and mini-programs in our 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 possible pair of user i and mini-program j, the dependent variable Y_{ij} equals 1 if the user authorizes data sharing with or initially visits the mini-program, and 0 otherwise. Like the user-level regression specified in Equation (1), Age_i, $Digital\ Experience_i$, and δ_i represent controls for user characteristics. Different from the user-level regression, this regression allows us to include mini-program fixed effects γ_j, which control for the heterogeneity across mini-programs.

Panel B of Table 3 reports the analysis at the user-mini–program level. Even after controlling for mini-program fixed effects, the results are very similar to those from the user-level analysis. Without and with the controls for user and mini-program characteristics, there is no significant difference in the number of data-sharing authorizations across "concerned," "very concerned," and "unconcerned" users, even though the level of privacy concerns is positively correlated with the propensity to have an initial visit to a mini-program.

We have also explored how data sharing authorizations may vary across users with different characteristics. In Panel C of Table 3, we expand the regression at the user-mini–program level specified in Equation (2) by interacting dummy variables $Concerned_i$ and $Very\ Concerned_i$ with other user characteristics. We focus on two characteristics: education and self-control. We define $Education_i$ as a dummy variable that indicates whether a user has a college degree or higher. We measure $Self\ Control_i$ by whether a user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the period from July 2019 to July 2020.[8] Interestingly, the first column shows that higher-educated users who are "concerned" or "very concerned" tend to authorize data sharing more than their "unconcerned" counterparts. This implies that the data privacy paradox is notably intensified among the well-educated. Conversely, the second column shows no significant variance between "concerned" and "very concerned" users compared to "unconcerned" ones in terms of the self-control metric. Thus, the data privacy paradox isn't exclusive to users with limited educational backgrounds or weaker self-control.

Overall, the results from Table 3 reject Hypothesis 1 and instead confirm the data privacy paradox that the respondents' data-sharing authorizations are not negatively related to their privacy concerns. This finding contradicts the common wisdom that privacy-concerned users are more reluctant to share personal data.

B. Validating Survey-Based Privacy Concerns

A common critique of the data privacy paradox is that it may simply reflect inconsistencies in how respondents express their privacy preferences, rather than a true 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 behavior.

To address this concern, we leverage administrative data to examine whether survey-based measures of privacy concerns correlate with concrete user actions to protect their data. Specifically, we analyze two behaviors: canceling previously authorized data sharing with mini-programs and modifying Alipay's default privacy settings. Conceptually, users who express stronger privacy concerns should be more likely to take these actions.

[8] We classify a mini-program as seemingly addictive if its description includes keywords such as "game," "lottery," or "red envelope."

Our analysis is again divided into user-level and user–mini-program-level regressions. For the user-level analysis, we apply the regression model from Equation (1), replacing the dependent variable with an indicator for whether a user revoked any data-sharing authorization between January 2013 and July 2020 or modified Alipay's default privacy settings between May 2017 and April 2020.[9] Notably, taking either action requires not only privacy concerns but also awareness of how to revoke data-sharing permissions or adjust Alipay's settings. 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, our regression includes controls for digital experience, age, gender, and city fixed effects.

Table 4, Panel A presents results from the user-level regressions. Columns (1) and (2) use the *'Has Canceled'* indicator as the dependent variable. Controlling for other factors, users who identified as *concerned* or *very concerned* about data sharing were significantly more likely to have revoked permissions for at least one mini-program compared to their *unconcerned* counterparts. Notably, the likelihood of cancellation was even higher among the *very concerned* group than the *concerned* group.

Columns (3) and (4) shift the focus to the *'Privacy Setting Changed'* indicator. Without controls, users with stronger privacy concerns were more likely to have modified Alipay's default settings compared to the *unconcerned* group. However, once extensive controls are introduced 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 are strongly associated with taking these privacy-protective actions, suggesting that younger and more digitally experienced users are better equipped to manage their privacy settings.

Panel B of Table 4 extends the analysis to the user–mini-program level, focusing on data-sharing cancellations. This granular approach allows us to control for mini-program-specific effects, offering insights into how users with different privacy concerns vary in their likelihood of revoking data-sharing agreements for the same mini-programs. Using the regression model from Equation (2), we analyze all active data-sharing consents between users and mini-programs from July 2019 to July 2020, with a sample size of 481,143. The results show that *'very concerned'* users exhibit a significantly higher tendency to cancel data-sharing authorizations.

[9] Alipay started to record these variables at different points in time, leading to their different periods of measurement.

Overall, Table 4 confirms that survey-based measures of privacy concerns align with actual privacy-protective actions taken by users, addressing concerns about the reliability of self-reported privacy attitudes. Our findings suggest that Solove's (2021) skepticism about the validity of privacy paradox studies does not apply here. Thus, the data privacy paradox persists even in the absence of measurement error in privacy concerns.

C. The Role of Omitted Variables

How can we explain the lack of a negative correlation between privacy concerns and the frequency of data-sharing authorizations? Building on the simple framework, this paradox can be reconciled if users who are more apprehensive about sharing personal data also perceive greater benefits from using mini-programs. If this holds, then the regressions specified in equations (1) and (2) are misspecified, as they fail to account for a user's underlying demand for digital services.

To address this, we extend the use-month regression specified in equation (1) by incorporating measures of the user's digital demand:

$$Y_{it} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + a_5 Demand_{it} + \delta_i + \lambda_t + \epsilon_i, \quad (3)$$

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. A session may involve multiple launches.

Table 5 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 5 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 results 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, higher 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 Demands

How does digital demand affect users' privacy concerns and data sharing? Extensive literature suggests that concerns about data privacy are not purely innate but are shaped through experience—emerging as users engage with digital applications and share personal data (Stigler, 1980; Posner, 1981). The cost of privacy, particularly when personal data is compromised (Fainmesser et al., 2019), increases with the volume of shared data. Extensive data sharing also enables digital providers to refine price discrimination strategies (Taylor, 2004; Acquisti and Varian, 2005) and exploit user vulnerabilities (Liu et al., 2020). In these scenarios, the more data consumers share, the more their privacy concerns are likely to intensify.

These mechanisms collectively suggest that greater digital engagement leads to stronger privacy concerns, as summarized in the following hypothesis:

Hypothesis 2: *Holding all else equal, more active users of mini-programs tend to have stronger concerns about data sharing and are more likely to revoke data-sharing authorizations..*

In this section, we examine this 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 in the survey responses. In Subsection B, using the same IV approach, we show that greater digital engagement also increases the likelihood of canceling data-sharing authorizations. These findings help establish the positive correlation between b_i and c_i, which is essential for resolving the paradox observed in the previous section.

In Subsection C, we further explore another important dimension: data sharing induced by intensified digital usage may, in turn, enhance digital applications, thereby fueling even greater digital demand. As consumers share more data, service providers leverage it to refine personalized services, making digital applications increasingly indispensable (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020; Cong et al., 2020).

This process operates as an equilibrium mechanism—rather than relying on data from a single individual, service providers optimize their platforms using data from a broad user base. As a result, users who were initially reluctant to share data may find their concerns offset—or even overridden—by the increasing utility of enhanced digital services. This dynamic suggests that data-sharing behavior may persist or even intensify over time, reinforcing or deepening the observed paradox. To investigate this further, in Subsection C, we examine whether privacy-concerned users decrease or increase their intensity of data-sharing authorizations post-survey.

A. Causal Effect of Digital Usage on Privacy Concerns

Although the digital economy is still in its early stages, Figure 2 reveals a rising trend in privacy concerns. The figure 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. While this figure is based on cross-sectional data, capturing users with varying durations on Alipay, the pattern suggests that deeper engagement with digital platforms is associated with heightened privacy concerns. This observation further underscores the need to investigate the causal relationship between digital usage and the evolution of privacy concerns.

The widespread adoption of Alipay in China is largely driven by its integration of various daily-life services, including a bike-sharing feature embedded in its mini-program ecosystem. In

the three years preceding our July 2020 survey, China's bike-sharing sector experienced significant growth, with multiple companies partnering with Alipay to enable users to scan and unlock bikes via the app. As these companies expanded their fleets and distributed bikes across more cities, local residents became increasingly reliant on Alipay—first to access shared bikes and subsequently to engage with other mini-programs. Since bike placements are determined by business and competitive strategies rather than individual privacy concerns, they provide a useful instrument for analyzing how digital engagement, exogenously influenced by bike placements, affects privacy concerns.

Ouyang (2022) employed a similar strategy to examine how mobile payment adoption, instrumented by placement of shared bikes bundled with Alipay, facilitates consumer credit. His analysis demonstrates that bike placement follows a staggered and largely exogenous pattern across cities, shaped by competitive pressures, logistical constraints, and capital availability.

To further strengthen the predictive power of bike placement, we further interact this measure with a user's prior experience with shared bikes. Intuitively, for users who have previously used shared bikes, an increase in bike availability is likely to spur even greater bike usage, leading to more frequent engagement with Alipay and its mini-programs. This instrumental variable, constructed as the interaction between prior bike usage and additional bike placements in a user's city, parallels the Bartik instrument widely used in the labor and trade literature. This instrument not only takes advantage of the staggered bike placement across cities but also the difference in each user's prior experience with shared bikes.

We employ a two-stage least squares (2SLS) framework and present the results of the IV analysis in Table 6. As before, we use three measures of a user's mini-program engagement: (1) the number of pages a user opened in mini-programs within a month, (2) the number of times relevant information was transmitted from servers while using mini-programs, and (3) the number of sessions in which the user engaged with mini-programs within a month. We take the logarithm of these measures in our analysis. We take the logarithm of these measures in our analysis.

Panel B reports the first-stage regressions of mini-program usage on the instrument—the interaction of $log(Bike\ Placement)_{ct}$ and $Have\ Started\ Using\ Bikes_{it}$. Columns (1)–(2) use $log(1+\#Visited\ Pages)_{it}$ to measure mini-program engagement, columns (3)–(4) use $log(1+\#\ App\ Launches)_{it}$, and columns (5)–(6) rely on $log(1+\#\ App\ Uses)_{it}$. We control for city, gender, and time fixed effects and cluster standard errors at the individual level to account for repeated

observations of the same individuals over time. The first-stage results indicate that the instrument is strong, with F-statistics ranging from approximately 29 to 778 across different regression specifications and mini-program usage measures, confirming that the instrument reliably predicts mini-program usage within our sample.

A valid instrument must also satisfy the exclusion restriction, meaning that in this context, the placement of shared bikes should affect privacy concerns only through increased mini-program usage, with no other direct pathways. It is evident that there is no direct mechanism linking bike placement to individuals' privacy concerns—simply seeing more bikes on the street does not inherently alter concerns about data sharing with Alipay's mini-programs. Any effect must stem from actively using shared bikes, which in turn increases engagement with Alipay and its mini-program ecosystem.

One potential concern is that a user's prior experience with shared bikes could be correlated with their privacy preferences. However, the data-sharing consent required for prior shared bike usage suggests that these users were likely less concerned about data privacy. This makes it even less likely that the instrument influences users' privacy concerns through any direct channel, further strengthening its validity.

Panel A of Table 6 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 either "concerned" or "very concerned" about data sharing with mini-programs. For the endogenous regressor, we use the fitted values of users' mini-program usage from the first-stage regressions, as shown in Panel B.

The second-stage coefficients consistently indicate a positive relationship between instrumented mini-program usage and privacy concerns. For instance, in columns (1)–(2), a 1% increase in the fitted number of visited pages corresponds to approximately a 0.038 percentage point increase in the probability of a user expressing privacy concerns. Columns (3)–(4) show an even larger coefficient, around 0.066, for app launches, while columns (5)–(6) indicate an effect size of about 0.145 for app uses. These magnitudes suggest that users who engage more deeply with Alipay's digital ecosystem, even when nudged exogenously, also become notably more concerned about data sharing. These effects are both statistically and economically significant, supporting Hypothesis 2 that greater immersion in data-driven digital apps can amplify anxieties about data sharing.

Although most of these estimates are highly significant at conventional confidence levels, some specifications—particularly those in columns (2), (4), and (6) with individual-level clustering—yield t-statistics ranging from 1.61 to 1.65, slightly above the typical 10% significance threshold. This modest drop in significance likely stems from the limited sample size of survey respondents and the stricter clustering approach, which inflates standard errors. However, in subsequent robustness checks using a larger and more representative sample, the same relationships remain strongly significant even under individual-level clustering, alleviating concerns about limited statistical power in the survey-based dataset.

By contrast, Panel C presents the OLS results from regressing privacy concerns on actual mini-program usage. The notable discrepancy between the OLS and 2SLS coefficients, with OLS estimates being considerably smaller, suggests that the OLS results may underestimate the true effect. This underestimation likely arises from selection bias, which the 2SLS approach effectively mitigates.

For instance, individuals who are less concerned about privacy may be more inclined to engage with mini-programs, causing the observed correlation between usage and privacy concerns to appear weaker than it truly is. Put differently, those who start with lower privacy concerns may already be "high-usage" users and remain less likely to express privacy worries, thereby attenuating the observed relationship. This comparison underscores the importance of using instrumental variables to identify causal effects, particularly in complex settings where direct observational relationships may be confounded by selection biases.

In summary, the IV analysis confirms that greater engagement with digital apps intensifies concerns about data privacy.

B. Causal Effect of Digital Usage on Data Sharing Cancellation

To further investigate whether increasing digital engagement heightens privacy concerns, we examine whether users' monthly interactions with mini-programs in Alipay increase the likelihood of actively revoking previously granted data-sharing authorizations. This test provides an even stronger measure than self-reported privacy concerns in surveys, as one might intuitively expect that more active users—those who derive the most benefit from mini-program services—would be less inclined to cancel their permissions. However, we posit that as users' digital usage increases,

they may become more concerned by potential privacy risks, making them more likely to withdraw authorizations.

We test this hypothesis by focusing on $\#\,Newly\,Cancelled\,Authorizations_{it}$, which captures, month by month, how many authorizations a user cancels among the mini-programs they had previously allowed.

We again employ a Bartik-style instrumental variable approach to address endogeneity concerns in measuring digital usage, leveraging the staggered rollout of shared bikes bundled with Alipay. Specifically, our instrument is the interaction of $log(Bike\ Placement)_{ct}$ with $Have\ Used\ Bikes_{it}$, a user-level indicator for whether user i has used Alipay-bundled shared bikes in month t. This variable differs from the broader indicator of whether a user has ever used shared bikes, as it is restricted to the specific month. This refinement allows us to capture more granular variation in bike-riding behavior across months, providing a richer perspective on how monthly expansions in city-level bike availability influence user activity. We expect that for users who actually ride the bikes in a given month, an increase in local bike supply will lead to a sharper rise in Alipay usage, which, in turn, may prompt more deliberate privacy-related actions—namely, canceling authorizations.

Table 7 presents the 2SLS regression results across three panels. In Panel B, the first-stage results confirm that our instrument strongly predicts the three measures of monthly mini-program usage. The F-statistics range from approximately 731 to 2,311, well above conventional weak identification thresholds.

Panel A presents the second-stage 2SLS estimates, linking exogenous changes in mini-program usage to monthly cancellations of data-sharing authorizations. Across all specifications, with or without individual-level clustering, the second-stage coefficients indicate a positive and statistically significant relationship between user activity and the number of newly canceled authorizations.

Economically, the estimated effects range from 0.013 to 0.031, implying that a 1% increase in usage corresponds to an approximately 1.3 to 3.1 basis-point rise in the number of canceled authorizations each month. This finding challenges the intuitive expectation that more active users—who benefit the most from mini-program services—would be less likely to restrict data flows. Instead, it suggests that higher engagement leads users to revoke sharing permissions more frequently, potentially reflecting heightened concerns about their data privacy.

Panel C of Table 7 presents the OLS estimates for comparison. Interestingly, unlike the privacy-concern regressions in Table 6, the OLS and 2SLS results here appear quite similar. This similarity may stem from the more directly observable nature of cancellation events, as well as potentially lower measurement error in tracking monthly revocations. Nonetheless, the 2SLS approach remains crucial for ensuring that we isolate exogenous variation in usage rather than conflating it with unobserved factors—such as a user's baseline propensity for risk-taking or impulsiveness—that could otherwise bias a purely observational regression.

Taken together, our findings from Tables 6 and 7 confirm Hypothesis 2, demonstrating that more active use of digital applications can lead users to become increasingly concerned about data privacy and more likely to revoke data-sharing authorizations.

Several studies have employed experimental approaches to estimate how much consumers value their data privacy. Acquisti, John, and Lowenstein (2013) use a field experiment to show that consumers' privacy valuations are highly sensitive to contextual and non-normative factors. Tang (2020) leverages a natural experiment in fintech loan applications, which require applicants to disclose certain personal information, to study how privacy concerns affect consumer behavior. Lin (2022) designs an experiment to distinguish between two types of privacy preferences: instrumental privacy preferences, which respond to the economic implications of revealing one's type, and intrinsic privacy preferences, which remain independent of economic payoffs. Our analysis builds on these insights by suggesting that privacy concerns are not only context-dependent but also cumulative—growing as users engage more deeply with data-sharing platforms. This pattern aligns with instrumental privacy preferences, where concerns about data usage evolve as users gain greater exposure to digital ecosystems.

C. The Evolution of Data Sharing Over Time

With the deepening of the digital economy, consumers are becoming increasingly engaged in digital applications. Our findings—that privacy concerns grow with digital usage—might suggest that users would become more reluctant to share their data over time. In this subsection, we examine whether the heightened privacy concerns established in our earlier analysis lead to a decline in data-sharing authorizations over time. Given the limited time span of our dataset, we focus on changes in data-sharing authorizations between the pre- and post-survey periods.

Panel B of Figure 1 summarizes data-sharing activity and initial visits during the post-survey period (August 2020 to December 2021). On average, "unconcerned" users engaged with 27.8 mini-programs and shared data with 22.5, while "concerned" users engaged with 32.8 and shared data with 24.6. Similarly, "very concerned" users engaged with 33.4 mini-programs and shared data with 23.8. A clear trend emerges: all groups exhibited increased activity and data-sharing post-survey compared to pre-survey, with "concerned" and "very concerned" users outpacing their "unconcerned" counterparts in data-sharing growth.

In Table 8, we formally examine changes in data-sharing authorizations across the three groups, expanding the user-mini-program-level regression specified in Equation (2). We extend the dataset to cover both pre- and post-survey periods, allowing us to test changes across time. The results indicate that "concerned" and "very concerned" users were significantly more likely than "unconcerned" users to authorize mini-programs in the post-survey period, even after controlling for user characteristics and mini-program fixed effects.

Both Figure 1 and Table 8 highlight the same pattern—users authorized more data-sharing post-survey, with the increase being particularly strong among those who had expressed privacy concerns in the survey. This evolution over time may seem counterintuitive, as it appears to contradict our earlier finding that increased digital usage heightens privacy concerns. However, this pattern reflects the data privacy paradox discussed in the previous section, where users' stated concerns about privacy do not necessarily translate into restrictive data-sharing behaviors.

As we demonstrated earlier, the data privacy paradox ultimately stems from an omitted variable problem—failing to account for the role of digital demand when examining the relationship between data sharing and privacy concerns. A similar omission may also explain the observed evolution of data-sharing behavior over time. It is crucial to recognize that digital demand itself is changing, alongside the growth in privacy concerns.

Over time, digital service providers, leveraging increasingly extensive and granular user data, have been able to develop more powerful, personalized, and convenient digital services tailored to individual users. These enhanced services likely fuel users' growing digital demand, which may offset or even override their rising privacy concerns. While our dataset does not allow us to directly test the extent of improvements in mini-program services on Alipay, our finding that users authorized more data sharing over time ultimately reflects their increasing digital demand, reinforcing the idea that privacy concerns alone do not dictate user behavior in digital ecosystems.

This straightforward yet encouraging finding—that data-sharing authorizations continue to grow despite rising privacy concerns—reinforces a pivotal premise for the digital economy: as it expands, it has the potential to offer increasingly powerful and personalized digital services, which can, in turn, incentivize users to sustain or even increase data sharing, despite their heightened concerns about privacy risks.

V. A Representative Sample

Our survey sample tends to include more-active users, as they are more likely to complete the survey. This bias raises a natural concern about the generalizability of our findings. To ensure robustness, we constructed a representative sample of 100,000 users, randomly drawn from all active Alipay users, and found results consistent with those from our survey sample, thereby reinforcing the validity and applicability of our main findings. Additionally, we leveraged this representative sample to examine the differential impact of a privacy awareness event on heavy versus light users of mini-programs. This analysis yields new evidence supporting Hypothesis 2, which posits that privacy concerns intensify with greater digital usage.

A. Robustness

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 A2 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*, 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.[10]

In Table 9, 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 3, Panel A. The results confirm that users with greater privacy concerns tend to authorize data sharing with more mini-programs. This relationship remains significant even after controlling for strong indicators of user knowledge, such as digital experience and age, as well as gender and city-specific effects, indicating 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 5. The results reaffirm that the paradox becomes less pronounced after accounting for different usage variables, as the coefficient on Privacy Setting Changed is 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 6. 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 6 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.

B. Heterogeneous Responses to a Privacy Related Incident

To further examine how privacy concerns vary among users with different levels of digital engagement, we analyzed a notable event. On January 3, 2018, Alipay released its Annual User Footprint Report within the app, 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 their explicit consent if they overlooked the

[10] 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.

setting. This design flaw was quickly exposed and widely discussed 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.

Despite these corrective actions, the incident significantly heightened public awareness of data privacy issues, prompting an increase in Alipay users revoking data-sharing authorizations with mini-programs, as shown in Figure A6. This event serves as a natural experiment, enabling us to assess how users with varying levels of digital service usage reacted to the privacy concerns triggered by the incident.

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

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

$$\begin{aligned} Daily\ Cancellation\ Dummy_{i,t} = \alpha_0 + \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{H,\tau} \cdot Heavy\ User_i \cdot \mathbb{I}(t=\tau) \\ + \beta_{H,6} \cdot Heavy\ User_i \cdot \mathbb{I}(t \geq 6) + \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{L,\tau} \cdot Light\ User_i \cdot \mathbb{I}(t=\tau) \\ + \beta_{L,6} \cdot Light\ User_i \cdot \mathbb{I}(t \geq 6) + \delta_i + \varepsilon_{i,t}, \end{aligned} \tag{4}$$

where t represents the number of days after the incident on January 3, 2018, $Daily\ Cancellation\ Dummy_{i,t}$ 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, and $\varepsilon_{i,t}$ is the error that varies across individuals and over time.

Since this event occurred before our survey sample, we use the random sample to estimate the regression specified in Equation (4). 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, demonstrating heightened privacy concerns through a greater likelihood of canceling data-sharing authorizations. However, this response is temporary, possibly due to

Alipay's swift corrective actions and the incident gradually fading from social media discussions. This finding remains robust when we directly test the differential response between heavy and light users in Panel A of Figure A7.

As before, one might argue that heavy users' greater propensity to cancel data sharing reflects their better familiarity with the authorization settings in the Alipay app rather than heightened privacy concerns triggered by the incident. To address this concern, we focus on a subsample of Alipay 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 remaining users were already knowledgeable about how to cancel data-sharing authorizations before the incident.

Panel B of Figure 4 presents the estimated $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients for this subsample. Although the behavioral gap between heavy and light users narrows, it remains significant, with heavy users still more likely to cancel data sharing with mini-programs. The reduced gap suggests that knowledge also plays an important role in explaining heavy users' higher cancellation rates. For this subsample, we also directly test the differential response between heavy and light users in Panel B of Figure A7. The difference is statistically significant on days 0, 2, and 3 following the incident.

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 demands become more concerned about data privacy following the incident.[11] This evidence reinforces the notion that concerns about data privacy are positively correlated with digital service usage.

VI. Data Sharing and Credit Line

In the rapidly expanding digital economy, data sharing has become an increasingly important source of valuable information for financial service providers to assess consumer credit and offer more convenient financial services. As FinTech and BigTech lenders integrate alternative data sources into their models, the role of data sharing in shaping consumer credit decisions has grown

[11] 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.

even more prominent. Berg et al. (2020) highlight the predictive power of simple digital footprints in assessing default risk and expanding credit access, demonstrating how nontraditional user data can enhance lending decisions. Building on this insight, this section examines how data sharing with mini-programs on Alipay correlates with the credit lines users receive, using a representative sample from Ouyang (2022) that includes users' credit access data. This analysis further highlights the impact of digital engagement on credit access.

The dataset comprises panel data on 41,485 randomly selected Alipay users observed over a 41-month period from May 2017 to September 2020. Each user appears in an average of 32 months of observations. It integrates static user characteristics with time-varying measures, including 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 10 employs Poisson regressions to model the relationship between data sharing and credit lines, using a nonnegative dependent variable that often contains many zeros. Each column represents a different time horizon (t, t+1, t+2, t+3, and t+6), illustrating how cumulative data sharing in month ttt predicts a user's credit line in subsequent months. Specifically, we include $log(1 + Cumulative\ Authorized\ Mini\text{-}Programs)_{it}$ to indicate the number of mini-programs a user has authorized to access their data and $log(1 + Cumulative\ Cancelled\ Mini\text{-}Programs)_{it}$ as a proxy for data-sharing withdrawals. Additionally, we control for $log(1 + In\text{-}Person\ Digital\ Payment\ Amount)_{it}$, which captures a user's monthly in-person Alipay transaction volume. Including this control accounts for the relationship between credit lines and payment flows, as documented in Ouyang (2022).

Across all five columns, greater digital payment flow correlates with higher credit limits, reinforcing Berg et al. (2020)'s finding that richer digital footprints lead to more favorable lending terms. Similarly, a higher cumulative count of authorized mini-programs is positively associated with credit lines, suggesting that consistent data sharing reduces informational asymmetries and signals user creditworthiness.

Interestingly, the cumulative number of canceled mini-programs shows negative coefficients, implying that frequent withdrawals of data-sharing authorizations may raise lender concerns about transparency or stability. While these effects weaken over longer horizons, they remain economically and statistically significant up to month t+6.

Taken together, these findings suggest that users with expansive and uninterrupted digital footprints gain greater credit access, while those who frequently curtail or revoke data sharing may forgo some potential credit. These patterns highlight how data-sharing behavior shapes users' digital footprints, influencing lenders' decisions and credit provision.

VII. Conclusion

In this paper, we combine survey and administrative data to examine how Alipay users' data-sharing behavior with third-party mini-programs relates to their privacy concerns within a major financial platform. We find that privacy-concerned users authorize more—not less—data sharing than unconcerned users, confirming the data privacy paradox in a setting highly relevant to the booming digital economy.

Rather than attributing this paradox to measurement errors in survey responses, user resignation to privacy risks, or behavioral biases in data-sharing decisions, we identify an omitted variable problem. Our analysis reveals that privacy-concerned users also exhibit stronger demand for digital services, which may outweigh their privacy concerns, offering a new explanation for the paradox.

Furthermore, we highlight the joint dynamics of privacy concerns and digital demand in shaping data-sharing behavior. While privacy concerns grow with mini-program usage, so does the demand for digital services—ultimately leading to more data sharing over time, despite rising privacy concerns.

This dynamic reflects broader trends in data-driven innovations, including the rapid advancement of generative AI models like ChatGPT and similar technologies available through the GPT Store. As these services evolve, they deliver greater benefits but also heightened concerns over data privacy and security. Our findings underscore the complex trade-offs shaping the digital economy, with significant implications for both innovation and stability.

For financial firms, markets, and regulators, our study offers valuable insights. Despite rising privacy concerns, consumer demand for data-driven, personalized financial services is likely to grow, reinforcing the long-term viability of such innovations. However, privacy concerns could pose risks to financial stability and consumer trust if left unaddressed. Financial firms may need to strengthen privacy protections and invest in privacy-enhancing technologies to balance data utilization and user confidence. Regulators, meanwhile, face the challenge of protecting consumer

financial data while fostering innovation, ensuring that privacy safeguards do not stifle technological progress.

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.

Aridor, G., Che, Y. K., & Salz, T. (2020). The Economic Consequences of Data Privacy Regulation: Empirical Evidence From GDPR. National Bureau of Economic Research.

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., Buchak, G., & Gornall, W. (2022). Customer Data Access and Fintech Entry: Early Evidence from Open Banking. Social Science Research Network.

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. (2021). FinTech Lending (Working Paper 29421; Working Paper Series). National Bureau of Economic Research.

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.

Brandimarte, L., Acquisti, A. and Loewenstein, G., (2013). Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4(3), 340–347.

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. (2020). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science*, forthcoming.

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. (2019). Digital Privacy. Social Science Research Network.

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. (2020). 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. (2022). Data versus Collateral. Review of Finance, rfac022.

Hoffmann, C. P., Lutz, C., & Ranzini, G. (2016). Privacy Cynicism: A New Approach to the Privacy Paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(4).

Johnson, G., Shriver, S., & Goldberg, S. (2019). Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR.

Goldfarb, A., & Tucker, C. (2012). Shifts in Privacy Concerns. *American Economic Review,* 102(3), 349–53.

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.

Kesan, J. P., Hayes, C. M., & Bashir, M. N. (2015). A Comprehensive Empirical Study of Data Privacy, Trust, and Consumer Autonomy. *Indiana Law Journal*, 91, 267–352.

Lin, T. (2022). Valuing Intrinsic and Instrumental Preferences for Privacy. *Marketing Science*, forthcoming.

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.

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. (2020). The Market for Data Privacy (SSRN Scholarly Paper 3352175).

Rossi, A. G., & Utkus, S. P. (2020). Who Benefits from Robo-advising? Evidence from Machine Learning. SSRN Electronic Journal.

Sockin, M. & Xiong, W. (2022). Decentralization Through Tokenization. *Journal of Finance*, forthcoming.

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: The Data Privacy Paradox

This figure depicts the numbers of initial visits and 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?*" Panel A covers the pre-survey period from July 2019 through July 2020, while Panel B covers the post-survey period from August 2020 to December 2021.

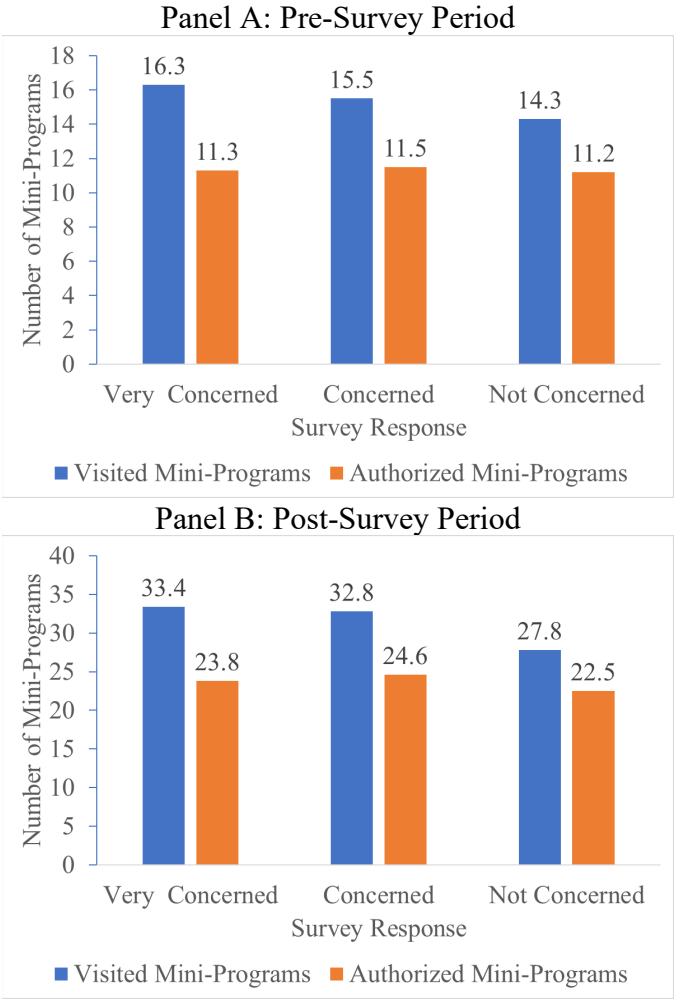


Figure 2: Digital Experience and Privacy Concerns

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.

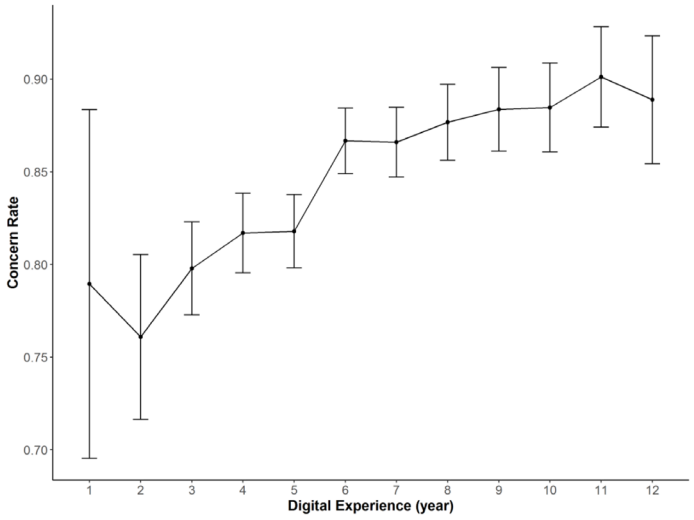
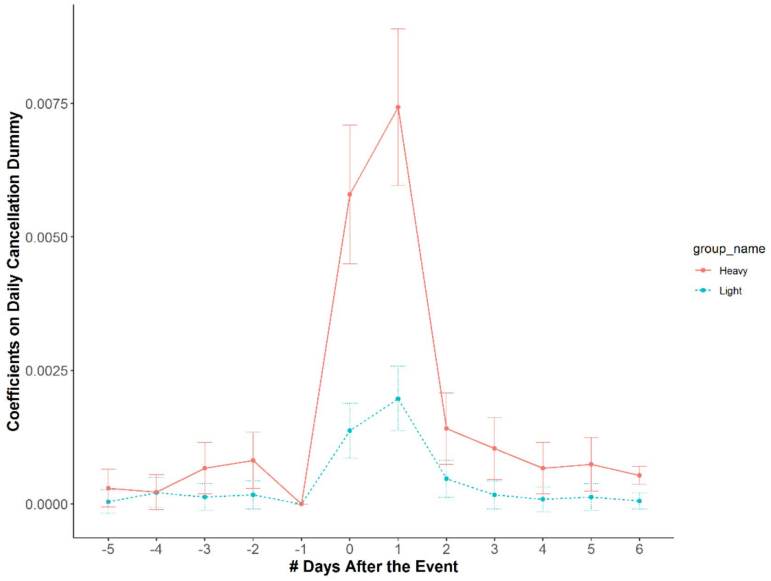


Figure 3: Activeness and Response to the 2017 Footprint Report Incident

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

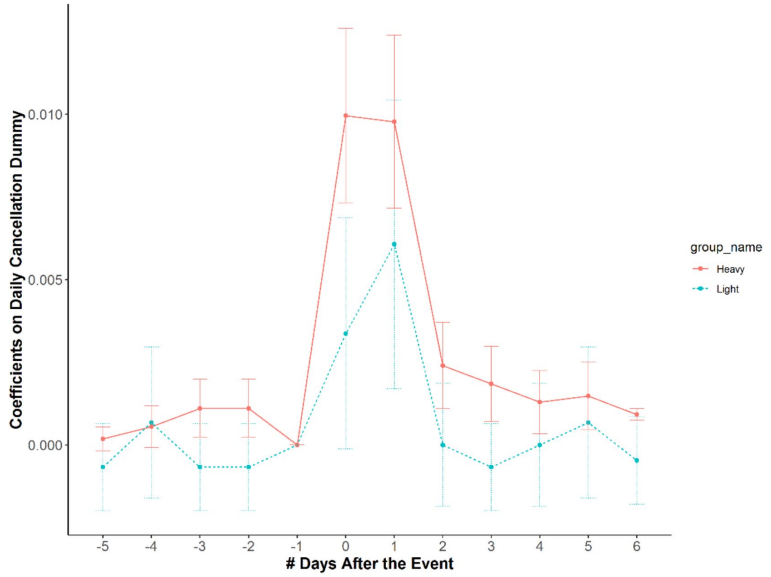


Table 1: Responses to Selected Survey Questions

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* and *Very Concerned Dummy* 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*, 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 entered 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* status, *# Cancellations,* and *Cancellation Rate* 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 active days, 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* equals 1 if a user is female, and 0 otherwise. *Education Dummy* 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
# Entered 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
Cancellation Rate$_i$	10,857	0.05	0.10	0.00	0.00	0.00	0.06	1.00
Monthly mini-program use								
# Active Days$_{it}$	1,521,645	0.57	2.92	0.00	0.00	0.00	0.00	31.00
# Uses$_{it}$	1,521,645	0.81	5.01	0.00	0.00	0.00	0.00	75.00
# 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	Concerned	Very Concerned	Difference	Difference
	(1)	(2)	(3)	(2) – (1)	(3) – (1)
Mean Digital Experience	66.868	75.725	76.961	8.857***	10.093***
				(1.018)	(0.996)
Mean Age	32.873	32.731	32.881	-0.142	0.008
				(0.300)	(0.293)
Mean Female Dummy	0.148	0.282	0.280	0.134***	0.132***
				(0.013)	(0.012)
Mean Education Dummy	0.137	0.221	0.214	0.084***	0.077***
				(0.012)	(0.012)

Table 3: The Data Privacy Paradox

This table presents regression analysis of the data privacy paradox results for the pre-survey period from July 2019 through July 2020. *Concerned Dummy* and *Very Concerned Dummy* in 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* is a dummy indicating whether the user has a college degree or higher. *Self Control* 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 and columns (3)–(4) for the monthly number of initially visited mini-programs. Panel B provides results for regressions at the user-mini–program level. In each pair of user-mini–program, columns (1)–(2) indicate results for the dummy whether the user allowed the authorization and columns (3)–(4) for the dummy whether the user visited at least once. Panel C reports heterogeneity analysis of the data privacy paradox, where we interact privacy concern measures with a user's personal characteristics. 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*$_{it}$	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$	0.023	0.014	0.089***	0.088***
	(0.015)	(0.015)	(0.023)	(0.023)
Very Concerned Dummy$_i$	0.008	-0.001	0.142***	0.140***
	(0.015)	(0.015)	(0.024)	(0.024)
Digital Experience$_{it}$		0.001***		-0.0002
		(0.0002)		(0.0003)
Age$_{it}$		-0.003***		0.015***
		(0.001)		(0.001)
Constant	0.800***		1.023***	
	(0.013)		(0.020)	
City, Gender, Time FE	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual
Observations	152,110	151,886	152,110	151,886
Adjusted R^2	0.00002	0.005	0.001	0.013

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$		*Visited Dummy*$_{ij}$	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ ($\times 10^{-4}$)	0.862	0.386	2.897***	2.552***
	(0.745)	(0.735)	(0.848)	(0.836)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	0.028	-0.465	3.755***	3.340***
	(0.736)	(0.728)	(0.846)	(0.840)
Constant	0.004***		0.005***	
	(0.0001)		(0.0001)	
Mini-program, City and Gender FE	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual
Observations	25,414,875	25,364,288	25,414,875	25,364,288
Adjusted R^2	0.000	0.105	0.000	0.129

Panel C: Heterogeneity Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$		*Visited Dummy$_{ij}$*	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ ($\times 10^{-4}$)	-0.649	0.600	1.592*	2.826***
	(0.811)	(0.771)	(0.931)	(0.875)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	-1.096	-0.327	2.696***	3.679***
	(0.807)	(0.765)	(0.939)	(0.881)
Concerned Dummy$_i$ × Characteristics Measure$_i$ ($\times 10^{-4}$)	5.120***	-2.644	4.354**	-3.283
	(1.833)	(1.855)	(2.001)	(2.096)
Very Concerned Dummy$_i$ × Characteristics Measure$_i$ ($\times 10^{-4}$)	3.329*	-2.501	2.982	-3.619*
	(1.800)	(1.827)	(1.981)	(2.083)
Characteristics Measure$_i$	-0.000	0.001***	0.000	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Characteristics Measure	Education	Self-Control	Education	Self-Control
Mini-program, City and Gender FE	Y	Y	Y	Y
Control Age and Digital Experience	Y	Y	Y	Y
Cluster	Individual	Individual	Individual	Individual
Observations	25,364,288	25,364,288	25,364,288	25,364,288
Adjusted R^2	0.105	0.105	0.129	0.129

Table 4: 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* and *Very Concerned Dummy* 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 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 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 July 2019 through July 2020. 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$_i$ (0/1)		*Privacy Setting Changed$_i$ (0/1)*	
	(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 and Gender FE	N	Y	N	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)
Concerned Dummy$_i$	-0.001	0.004
	(0.003)	(0.003)
Very Concerned Dummy$_i$	0.005	0.011***
	(0.003)	(0.003)
Digital Experience$_i$ ($\times 10^{-4}$)		1.218***
		(0.305)
Age$_i$ ($\times 10^{-4}$)		2.547**
		(1.141)
Constant	0.058***	
	(0.003)	
Mini-program, City and Gender FE	N	Y
Observations	481,143	480,542
Adjusted R^2	0.0001	0.107

Table 5: The Data Privacy Paradox Gone after Controlling Digital Usage

This table shows that after adding digital usage as controls, the data privacy paradox gone. *Concerned Dummy* and *Very Concerned Dummy* 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 launches columns (3)–(4), and the number of 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. 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, Time FE	N	Y	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual	Individual	Individual
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$_i$	-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$_i$	-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	Individual	Individual	Individual	Individual	Individual	Individual
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 6: Causal Effect of Digital Demand on Privacy Concerns

This tables examine the causal relationship between digital demand and revealed privacy concern in the survey. *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." *Have Started Using Bikes*$_{it}$ is one if a user *i* have started using Alipay-bundled shared bicycles before the month *t*, and zero otherwise. *log(Bike Placement)*$_{ct}$ indicates the log-transformed number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* in month *t.* We use the number of visited pages columns (1)–(2), the number of launches columns (3)–(4), and the number of uses columns (5)–(6) to capture the demand for digital services. This table 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 concerns. Panel A outlines the 2SLS estimates; Panel B describes the first stage; Panel C 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 A: Two-Stage Least Squares						
$\log(1 + \#\ \widehat{\text{Visited Pages}})_{it}$	0.038***	0.038				
	(0.007)	(0.023)				
$\log(1 + \#\ \widehat{\text{App Launches}})_{it}$			0.066***	0.066		
			(0.012)	(0.040)		
$\log(1 + \widehat{\#\ \text{App Uses}})_{it}$					0.145***	0.145
					(0.027)	(0.090)
Panel B: First Stage for Usage Variables						
$\log(\text{Bike Placement})_{ct} \times$ Have Started Using Bikes$_{it}$	0.028***	0.028***	0.017***	0.017***	0.008***	0.008***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
F-Statistic	778.19	180.38	396.19	82.27	143.38	29.12
Adjusted R^2	0.047	0.047	0.064	0.064	0.104	0.104
Panel C: Ordinary Least Squares						
log(1+# Visited Pages) $_{it}$	0.005***	0.005***				
	(0.0005)	(0.001)				
log(1+# App Launches) $_{it}$			0.007***	0.007***		
			(0.001)	(0.001)		
log(1+# App Uses) $_{it}$					0.009***	0.009***
					(0.001)	(0.002)
Adjusted R^2	0.043	0.043	0.044	0.044	0.044	0.044
City, Gender, Time FE	Y	Y	Y	Y	Y	Y
Cluster	N	Individual	N	Individual	N	Individual
Observations	147332	147332	147332	147332	147465	147465

Table 7: Digital Demand and Cancellation

This table examines the relationship between user activeness and cancellation of previously authorized mini-programs. The sample covers user-mini–program pairs that had been active between July 2019 and July 2020. *# Newly Cancelled Authorizations $_{it}$* refers to the number of cancelled authorizations of a user *i* in month *t*. *Have Used Bikes$_{it}$* is one if a user *i* have used Alipay-bundled shared bicycles in the month *t*, and zero otherwise. *log(Bike Placement)$_{ct}$* indicates the log-transformed number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* in month *t.* We use the number of visited pages columns (1)–(2), the number of launches columns (3)–(4), and the number of uses columns (5)–(6) to capture the demand for digital services. Panel A 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 cancellation. Panel B describes the first stage; Panel C provides the OLS results. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	# Newly Cancelled Authorizations $_{it}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Two-Stage Least Squares						
$\log(1+\#\widehat{\text{Visited Pages}})_{it}$	0.013**	0.013**				
	(0.006)	(0.006)				
$\log(1+\#\widehat{\text{App Launches}})_{it}$			0.019**	0.019**		
			(0.008)	(0.009)		
$\log(1+\widehat{\#\text{App Uses}})_{it}$					0.031**	0.031**
					(0.013)	(0.014)
Adjusted R^2	0.076	0.076	0.076	0.076	0.076	0.076
Panel B. First Stage for Usage Variables						
log(Bike Placement)$_{ct}$ × Have Used Bikes$_{it}$	0.077***	0.077***	0.053***	0.053***	0.033***	0.033***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
F-Statistic	2311.87	1272.55	1760.24	1078.30	1151.84	731.82
Adjusted R^2	0.296	0.296	0.344	0.344	0.376	0.376
Panel C. Ordinary Least Squares						
log(1+# Visited Pages) $_{it}$	0.015***	0.015***				
	(0.001)	(0.001)				
log(1+# App Launches) $_{it}$			0.020***	0.020***		
			(0.001)	(0.001)		
log(1+# App Uses) $_{it}$					0.026***	0.026***
					(0.001)	(0.002)
Adjusted R^2	0.076	0.076	0.076	0.076	0.076	0.076
Individual, Time FE	Y	Y	Y	Y	Y	Y
Cluster	N	Individual	N	Individual	N	Individual
Observations	147497	147497	147497	147497	147630	147630

Table 8: The Data Privacy Paradox in the Post-Survey Period

This table presents 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* and *Very Concerned Dummy* in 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." The results for regressions are at the user-mini–program level. In each pair of user-mini–program, columns (1)–(2) indicate results for the dummy whether the user allowed the authorization and columns (3)–(4) for the dummy whether the user visited at least once. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Authorized Dummy$_{ij}$ (0/1)		*Visited Dummy$_{ij}$ (0/1)*	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ ($\times 10^{-4}$)				
× Post-Survey Dummy$_{ij}$	274.587***	256.997***	3.314***	-9.945
	(53.601)	(52.630)	(0.530)	(9.511)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)				
× Post-Survey Dummy$_{ij}$	535.489***	502.886***	2.696***	-23.389**
	(53.176)	(52.183)	(0.524)	(9.492)
Concerned Dummy$_i$ ($\times 10^{-4}$)	-272.477***	-255.647***	-0.000	12.561
	(53.697)	(52.690)	(0.000)	(9.630)
Very Concerned Dummy$_i$ ($\times 10^{-4}$)	-534.387***	-502.433***	-0.000	25.502***
	(53.268)	(52.240)	(0.000)	(9.620)
Post-Survey Dummy$_{ij}$	-0.863***	-0.799***	-0.997***	-0.923***
	(0.004)	(0.004)	(0.000)	(0.001)
Constant	0.866***		1.000***	
	(0.004)		(0.000)	
Mini-program, City and Gender FE	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual
Observations	72,525,375	72,401,144	72,525,375	72,401,144
Adjusted R^2	0.298	0.367	0.368	0.440

Table 9: 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) and (2) show results for the number of authorized mini-programs, and columns (3) and (4) show results for the number of initially visited mini-programs. In columns (2) and (4), 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 launches columns (3)–(4), and the number of uses columns (5)–(6). Columns (1), (3), (5), and (7) show regression results without any other controls, while columns (2), (4), (6), and (8) control for digital experience and age, as well as user gender, user city, and time fixed effects. Panel C and Panel D examine 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 user level. *Have Started Using Bikes*$_{it}$ is one if a user *i* have started using Alipay-bundled shared bicycles before the month *t*, and zero otherwise. *log(Bike Placement)*$_{ct}$ indicates the log-transformed number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* in month *t.* We use the number of visited pages columns (1)–(2), the number of launches columns (3)–(4), and the number of 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 seeting. 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$_{it}$	
	(1)	(2)	(3)	(4)
Privacy Setting Changed$_i$	0.158***	0.136***	0.020***	0.175***
	(0.005)	(0.005)	(0.006)	(0.006)
Constant	0.195***		0.239***	
	(0.001)		(0.001)	
City, Gender, Time FE	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual
Observations	1,776,150	1,738,656	1,776,150	1,738,656
Adjusted R^2	0.004	0.020	0.004	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, Time FE	N	Y	N	Y	N	Y
Control Age and Digital Experience	N	Y	N	Y	N	Y
Cluster	Individual	Individual	Individual	Individual	Individual	Individual
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						
$\log(1 + \#\ \widehat{\text{Visited Pages}})_{it}$	0.140***	0.140***				
	(0.001)	(0.006)				
$\log(1 + \#\ \widehat{\text{App Launches}})_{it}$			0.206***	0.206***		
			(0.002)	(0.009)		
$\log(1 + \widehat{\#\ \text{App Uses}})_{it}$					0.394***	0.394***
					(0.004)	(0.016)
Panel C2. First Stage for Usage Variables						
$\log(\text{Bike Placement})_{ct} \times$ Have Started Using Bikes$_{it}$	0.039***	0.039***	0.027***	0.027***	0.014***	0.014***
	(0.0002)	(0.0004)	(0.0001)	(0.0003)	(0.0001)	(0.0002)
F-Statistic	52,113.18	8,628.95	43,583.51	7,221.53	30,802.59	4,822.24
Adjusted R^2	0.041	0.041	0.039	0.039	0.037	0.037
Panel C3. Ordinary Least Squares						
$\log(1+\#\ \text{Visited Pages})_{it}$	0.018***	0.018***				
	(0.0002)	(0.001)				
$\log(1+\#\ \text{App Launches})_{it}$			0.024***	0.024***		
			(0.0003)	(0.001)		
$\log(1+\#\ \text{App Uses})_{it}$					0.036***	0.036***
					(0.0004)	(0.001)
Adjusted R^2	0.010	0.010	0.010	0.010	0.010	0.010
City, Gender, Time FE	Y	Y	Y	Y	Y	Y
Cluster	N	Individual	N	Individual	N	Individual
Observations	1,721,344	1,721,344	1,721,344	1,721,344	1,785,600	1,785,600

Table 10: Digital Footprints, Data Sharing, and Financial Credit Access

This table examines the relationship between mini-programs data sharing with financial credit access. 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, *Cumulative Number of Authorized Mini-Programs*, indicates how many data a user shared. The second one, *Cumulative Number of Cancelled Mini-Programs*, is a proxy of withdraw from data sharing. The results for regressions are at the user-month level. We use credit line in month *t*, *t+1*, *t+2*, *t+3*, *t+6* 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+1*	*t+2*	*t+3*	*t+6*
	(1)	(2)	(3)	(4)	(5)
log(1+In-Person Digital Payment Amount)$_{it}$	0.007***	0.008***	0.009***	0.009***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(1+Cumulative Number of Authorized Mini-Programs)$_{it}$	0.368***	0.336***	0.306***	0.277***	0.203***
	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)
log(1+Cumulative Number of Cancelled Mini-Programs)$_{it}$	-0.181***	-0.177***	-0.168***	-0.160***	-0.137***
	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)
Individual, Time FE	Y	Y	Y	Y	Y
Observations	843,286	828,157	800,999	774,047	694,357
Adjusted R^2	0.867	0.867	0.869	0.872	0.883

The Data Privacy Paradox 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 A1: 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.

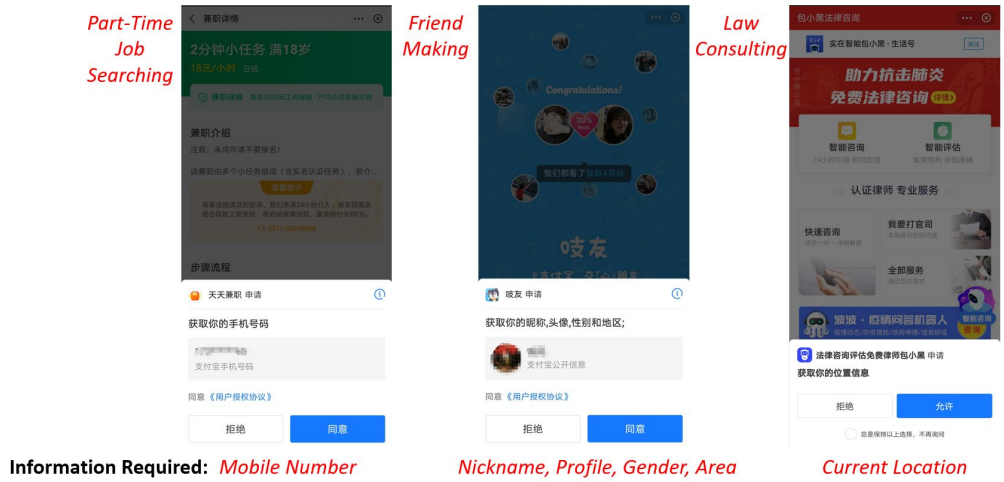


Figure A2: Completion Time Distribution (Seconds)

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.

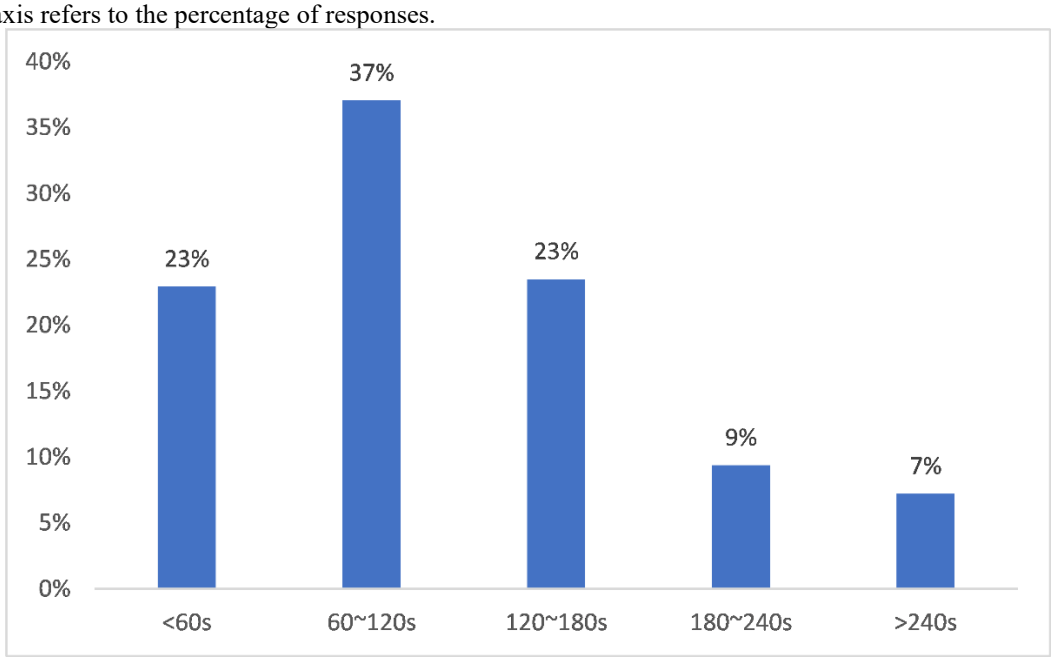


Figure A3: Age Distribution (Years)

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.

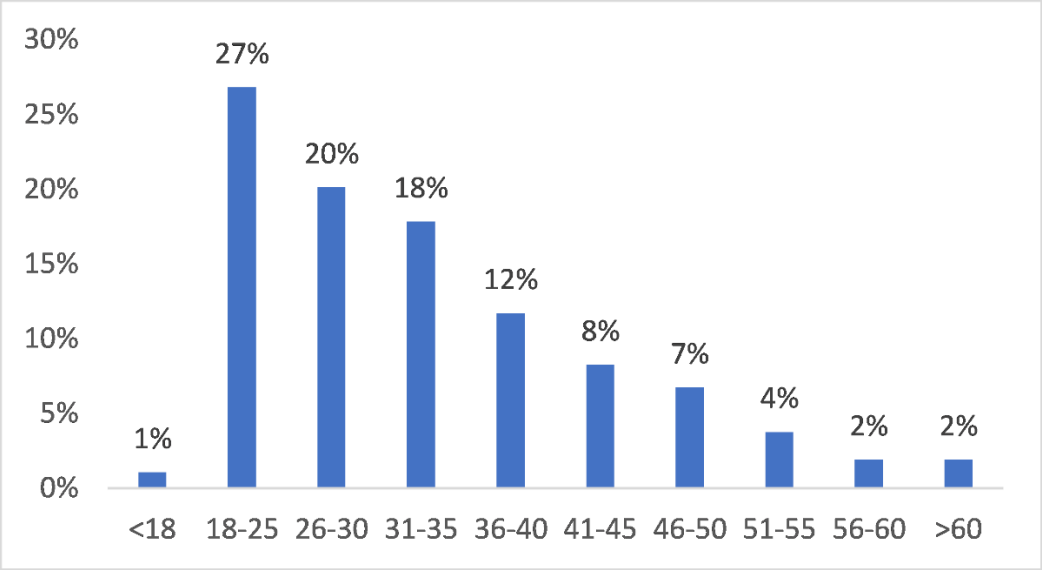


Figure A4: Distribution Across Provinces

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).

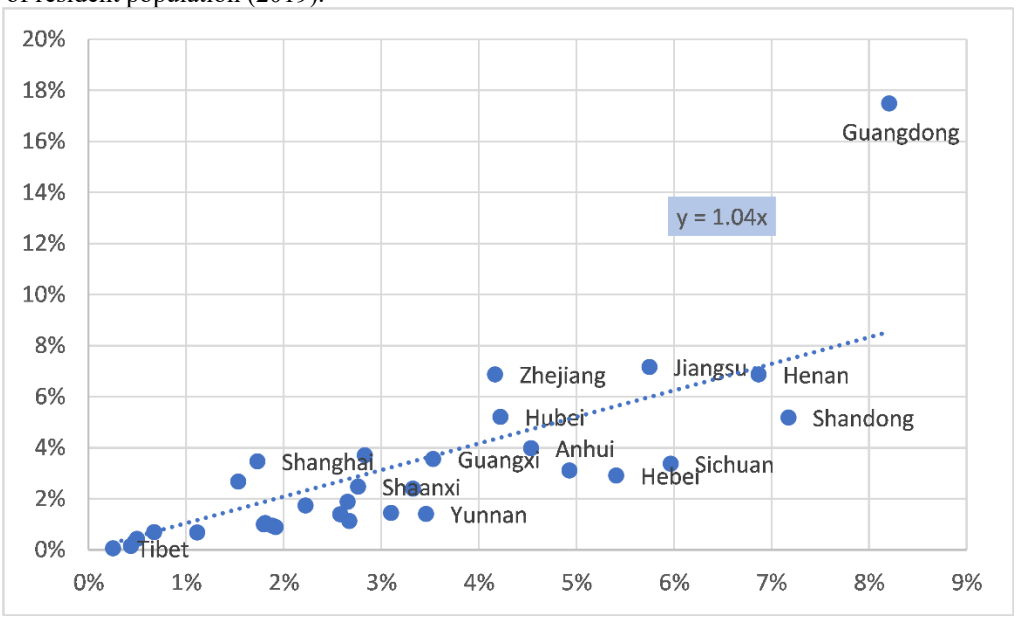


Figure A5: Digital Age Distribution (Years)

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.

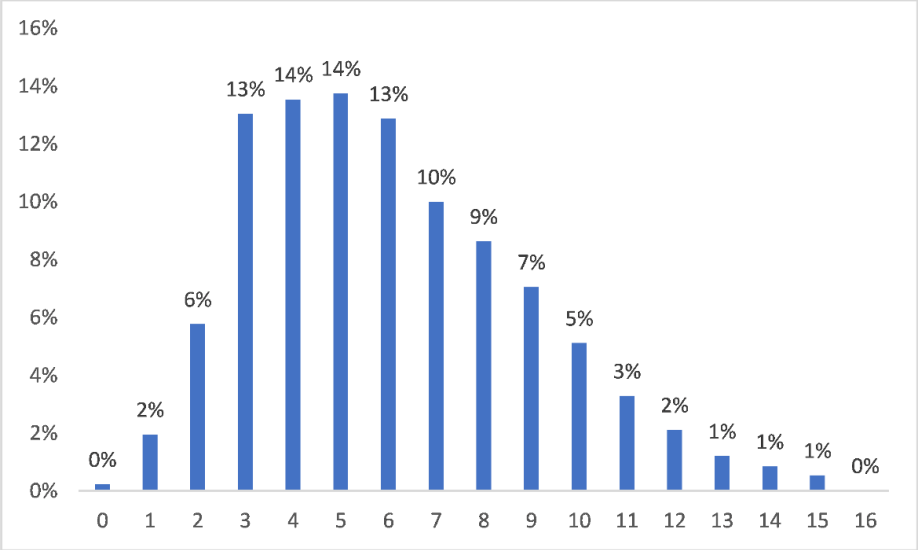


Figure A6: Cancellations During the 2017 Footprint Report Incident

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.

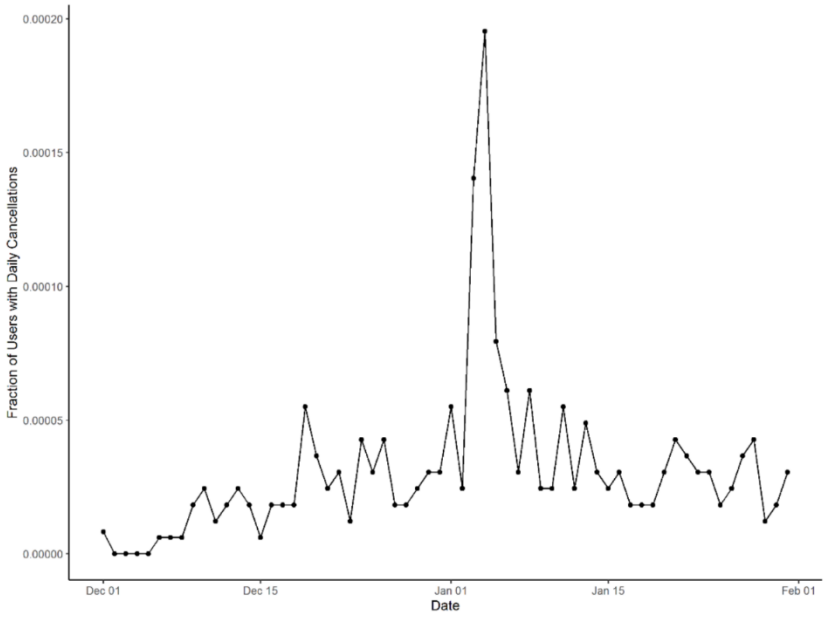


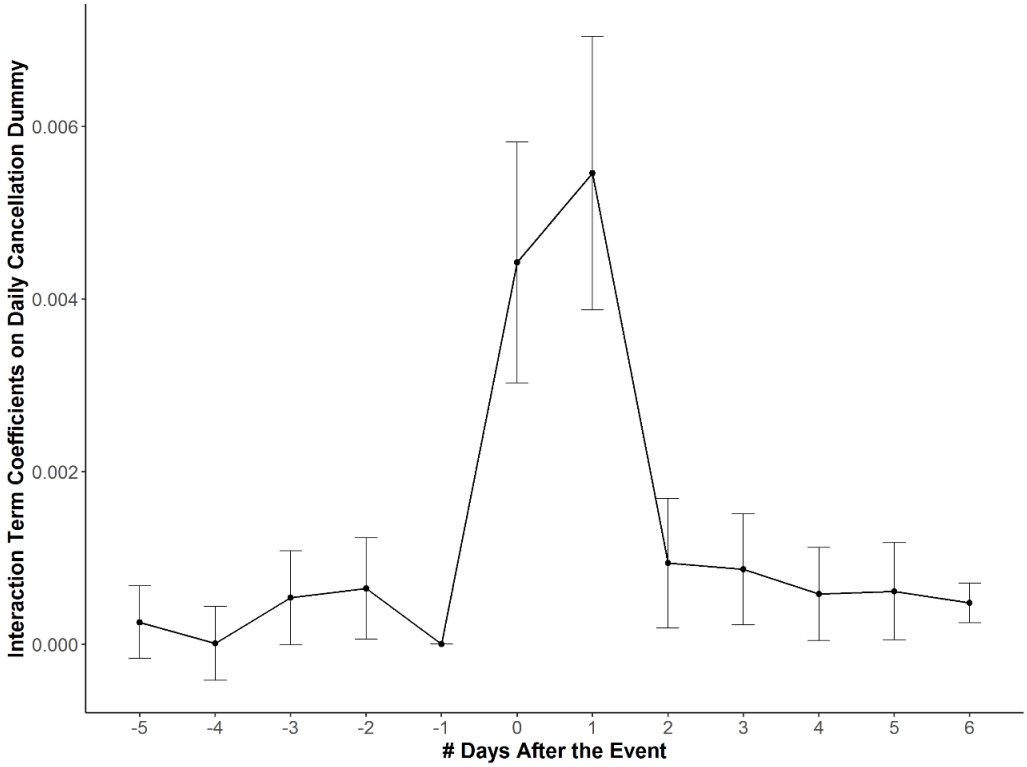
Figure A7: Activeness and Response to the 2017 Footprint Report Incident

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

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

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

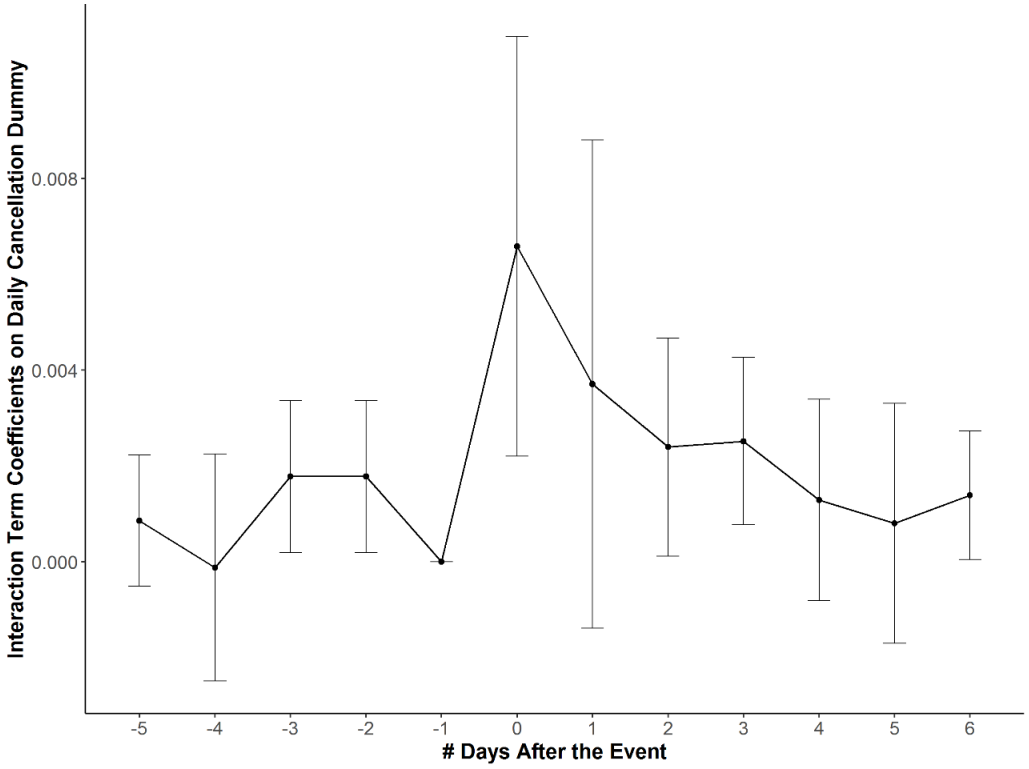


Figure A8: Time Trend in Demand for Digital Services

This figure shows the monthly time series of the average number of visited pages per active mini-program of Alipay users in three groups based on their responses to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" The sample covers the pre-survey period and focuses on the mini-programs that each user uses at least once during the corresponding months.

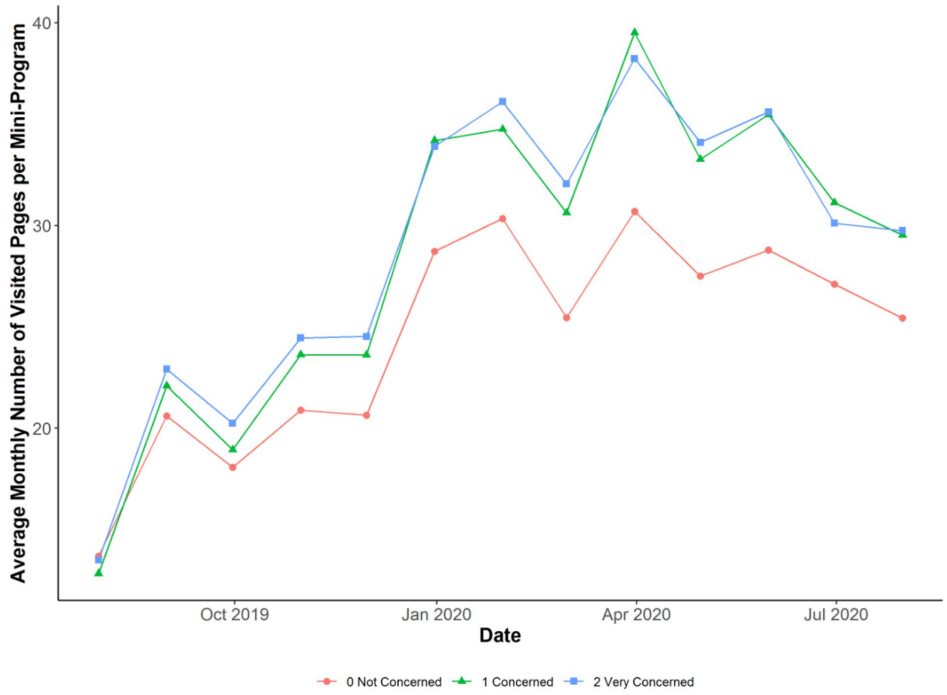


Table A1: Determinants of Data-Sharing Authorizations in Survey

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 and 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 A2: Responses to Other Selected Survey Questions

This table summarizes the responses of the respondents to five question in the survey related to privacy concerns, data sharing, and opting out from mini-programs. (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?*

	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 A3: Validating Selected Survey Responses

This table reports how the response in survey is related to privacy-seeking actions, including data-sharing authorizations with mini-programs, changing Alipay's default privacy settings, and canceling authorizations. *Authorization Rates in Survey* is average of the response from 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 that equal 1 if the answer to the survey question "*Have you ever changed your privacy settings in Alipay?*" is "yes". *Has Canceled in Survey* is a dummy variable that equal 1 if the answer to the survey question "*Have you ever opted out from mini-programs in Alipay?*" is "yes". *# Authorized Mini-Programs* is the number of authorized 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. *Privacy Setting Changed* is a dummy that indicates whether a user has changed Alipay's default privacy settings the period of May 2017 through April 2020. *Has Canceled* 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. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors 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$_i$		0.761***	
		(0.035)	
Has Canceled in Survey$_i$			0.798***
			(0.040)

Table A4: 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, entered, and canceled mini-programs over the pre-survey period of July 2019 to July 2020; the *Cancellation Rate* of used mini-programs between July 2019 and July 2020; and the *Has Canceled* status over the 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 active days, 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
# Entered 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
Cancellation Rate$_i$	98,674	0.01	0.07	0.00	0.00	0.00	0.00	1.00
Monthly mini-program use								
# Active Days$_{it}$	3,036,555	0.27	1.59	0.00	0.00	0.00	0.00	27.00
# Uses$_{it}$	3,036,555	0.34	2.21	0.00	0.00	0.00	0.00	40.00
# 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

Table A5: Heterogeneity Analysis of Data Privacy Paradox, by Gender and Age

This table presents heterogeneity analysis of the data privacy paradox results for the pre-survey period from July 2019 through July 2020, where we interact privacy concern measures with a user's personal characteristics. The personal characteristics is age, in which we take a value of 1 if a user's age is above the median of the sample. The personal characteristics is gender, in which we take a value of 1 if a user is female. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Authorized Dummy$_{ij}$		*Visited Dummy*$_{ij}$	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$	0.727	-0.119	2.256*	2.146**
	(0.920)	(0.831)	(0.985)	(0.940)
Very Concerned Dummy$_i$	-0.291	-0.429	2.827***	3.329***
	(0.903)	(0.823)	(0.976)	(0.947)
Concerned Dummy$_i$ × Characteristics Measure$_i$	-0.968	2.363	0.594	1.960
	(1.531)	(1.182)	(1.784)	(2.117)
Very Concerned Dummy$_i$ × Characteristics Measure$_i$	-0.496	0.445	1.182	0.560
	(1.511)	(1.781)	(1.780)	(2.102)
Characteristics Measure$_i$	-0.0002		-0.0003*	
	(0.0001)		(0.0002)	
Characteristics Measure	Age	Gender	Age	Gender
Mini-program, City and Gender FE	Y	Y	Y	Y
Control Age and Digital Experience	Y	Y	Y	Y
Cluster	Individual	Individual	Individual	Individual
Observations	25,364,288	25,364,288	25,364,288	25,364,288
Adjusted R^2	0.105	0.105	0.129	0.129