Transient Customer Response to Data Breaches of Their Information

Sumit Agarwal, Pulak Ghosh, Tianyue Ruan, and Yunqi Zhang*

August 2023

Accepted at *Management Science*

Abstract

Cybersecurity breaches pose a substantial concern in the digital era. We investigate how customers respond to multiple unexpected data breaches of their information in India. Difference-in-differences estimates show that digital payments declined by 9% relative to cash payments immediately after an unexpected data breach in a food delivery platform, but the gap disappeared three months later. Customer entry and exit also exhibit weak, short-lived changes. Additional analyses on bank and online grocery data breaches uncover even weaker effects of data breaches. Our findings imply that the perceived benefits of convenience outweigh the costs of payment security risks.

JEL codes: D12, D18, G29.

Keywords: cybersecurity breach, unexpected data breach, digital payment, cash payment, convenience, payment security, digital economy, India

---

* Sumit Agarwal (ushakri@yahoo.com) and Tianyue Ruan (tianyue.ruan@nus.edu.sg) are from the National University of Singapore, Pulak Ghosh is from the Indian Institue of Management Bangalore (pulakghoshc@gmail.com), and Yunqi Zhang (corresponding author) is from Nankai University (zhangyunqi@u.nus.edu). We thank David Sraer (the Editor), an anonymous Associate Editor, and four anonymous referees for their comments. We also thank Zefeng Chen, Yeqing Zhang, and participants at the Georgetown Global Virtual Seminar Series on Fintech, Asian Meeting of the Econometric Society 2022, China Fintech Research Conference 2022, and the 5th Big Data, AI and Fintech Conference for helpful comments and discussions. Agarwal acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 72150004). Ruan acknowledges financial support from the NUS Start-Up Grant No. A-0003870-00-00. Zhang acknowledges the financial support from the Ministry of Education in China Project of Humanities and Social Sciences (Project No. 20YJC790183) and the National Nature Science Foundation of China (Grant No. 72203112).

1. Introduction

Digital services, such as food delivery and ride-sharing, store and harvest private or even confidential consumer data. Given the massive volume of Internet activity, cybersecurity breaches can lead to large-scale data leakages and pose a substantial concern for users of digital services. Notable recent data breaches include Yahoo, Equifax, and eBay which affected 3 billion, 147 million, and 143 million accounts, respectively. It remains largely unknown how consumers respond to data breaches of their information: Do a breached digital service lose customers? How persistent is the effect? Does the data breach shift the composition of customers? We analyze consumer behaviors in an unexpected data breach to shed light on these open questions.

The data breach involved a leading online food delivery platform in India. The platform aggregates information on restaurants and allows consumers to order food and pay for the food delivery services by cash or digital payments (digital wallets and bank cards). The data breach was first reported by media on May 18, 2017 and involved 17 million users' names, email addresses and passwords; no payment information was leaked according to the firm. We use the customer receipt-level transaction data from this platform to analyze how the data breach affected consumer behaviors. The payment mode information allows us to examine the relative change between cash and digital payments after controlling for the time trends in a difference-in-differences (DID) framework. Digital payments are expected to decline relative to cash payments if consumers are concerned about their payment security.

Following the data breach, the overall usage exhibited a modest and temporary decline. Such a decline in overall usage was driven by digital payments; cash payments increased despite the data breach. Digital payments declined by 9% relative to cash payments during the first month after the breach. This divergence reveals that users worried about payment security even when no payment information was leaked. However, the decline was transient. The difference between digital and cash payments vanished in the third month after the breach.

To gauge whether the data breach changes customer composition, we also examine customer entries and exits. Both entry and exit intensities changed little following the data breach. The entrants after the data breach do not differ, in terms of observable characteristics, from the entrants before the data breach. Interestingly, compared to existing users who primarily used cash, those who primarily used digital payments were less likely to quit using the app following the data

breach. As the pre-breach usage of digital payments partly reflects demand for convenience, this result suggests that consumers who value convenience continue using the app but may switch from digital payments to cash payment, which is perceived to be safer.

Taken together, our analysis shows that the data breach indeed affects consumers in the short term and leads to a switch from digital payments to cash payments. In the long term, usages return to pre-breach levels. Compared to surveys that directly elicit consumer perceptions of data breaches, our approach of analyzing consumer activity uses data with little measurement error and permits a more powerful test. A limitation is that we are unable to directly measure consumer perceptions from actual behaviors. We discuss several potential mechanisms through which the data breach has a transient effect on consumer behaviors. Our preferred interpretation is as follows: While concerns for payment security triggered by the privacy leakage lead to a decline in digital payments in the short term, the perceived benefits of convenience outweigh the costs of the breach in the long term; as a result, usages return to pre-breach level.

We also analyze the consumer response to two other unexpected data breaches, one on an online grocery store and the other one a bank, using administrative data from the breached firms. We find that these two data breaches have little effect on consumer behaviors. The stability of the results across the multiple breaches demonstrates that the observed transient response is not driven by app- or industry-specific idiosyncratic characteristics. Overall, the examined breaches have different causes and types of what is typically considered non-sensitive data involved. Similar to the data breaches we analyze, the majority of data breaches do not compromise financial or personally identifying information.[1] The transient impact of data breaches on consumer behaviors that we document in our analysis is conceptually applicable to other data breaches of non-sensitive data.

We contribute to an emerging literature on consumer response to data breaches. When some studies use survey data (e.g., Greene and Stavins, 2017), evidence based on transaction data starts to emerge. For example, Schuh and Stavins (2016) and Mikhed and Vogan (2018) study how data

[1] The U.S. Government Accountability Office (2007) report on consumer data breaches considers "sensitive personal information" as financial information (bank account, credit card, financial investments, financial transactions) or personal identification information (Social Security numbers, government issued IDs). According to this definition, only 42 or 7.3% of the 577 data breaches that took place during the 11 years from 2011 to 2021 as compiled by Internet security service Have I Been Pwned (https://haveibeenpwned.com/) involved leakage of some sensitive personal information; the other 535 or 92.7% of the breaches did not involve leakage of any sensitive personal information

breaches of bank account affect consumers' card usage. Janakiraman, Lim, and Rishika (2018) studies how data breaches of a retailer affect their customers purchase from of the retailer. Our paper differs from the literature as the data breach of the food delivery platform only leaked platform login information and *did not* involve digital payment account information. Our data enable us to seperately analyze the effects on platform use and digital payment adoption.

We also contribute to the empirical literature on the digital economy and private information sharing (Argente, Hsieh, and Lee, 2022; Tang, 2019). Most existing studies discuss convenience and private information security separately (e.g., Gu et al., 2017; Rysman and Schuh, 2017; Teo et al., 2015). A few exceptions including Pentina et al. (2016) and Wottrich, van Reijmersdal, and Smit (2018) use survey-based or experiment data and find that the perceived information security concern does not prevent the adoption or future use of private-information sensitive apps. Our paper shows that the digital payment declines in the short run but gradually recovers as time goes on, which support the consumers' trade-off between payment security and convenience.[2]

Our findings shed light on the welfare implications of data ownership. Recent theoretical studies highlight non-rivalry and increasing returns to scale of data sharing; as a result, even in the presence of privacy considerations, consumer data sharing leads to social gains (Jones and Tonetti, 2020) and contributes to economic growth through R&D and knowledge accumulation (Cong, Xie, and Zhang, 2021). By indirectly showing that consumers are willing to share data, which can expose them to information security risk, in exchange for convenience, our analyses empirically corroborate these theoretical predictions.

2. Empirical Setting and Data Sources

We study how consumers respond to information leakage in an unexpected data breach. The breach occurred on a leading food delivery platform in India. The firm aggregates information on restaurants on an online platform and provides food ordering and delivery services from partner restaurants in select cities. On May 17, 2017, customer account information on the platform was discovered to have been sold on a dark web marketplace. On May 18, 2017, the firm confirmed

[2] Note the reversal in the effects we find also differs from the findings in the data breach literature. Greene and Stavins (2017) examine consumer rating on security, but only explore payment use in the long run. Schuh and Stavins (2016) explore in the short run, Mikhed and Vogan (2018) find no change in card use, Janakiraman, Lim, and Rishika (2018) find significant results for both short run and the long run. Kwon and Johnson (2015) find long run effects only.

this data breach, which was widely reported thereafter. The data breach involved 17 million users' names, email addresses, and passwords. The firm claimed that it had reset passwords for all affected users and logged them out of its app and website, and no payment information was leaked.

We use administrative data from the breached platform for our analysis. The dataset contains micro-level information about food ordering and delivery from April 2016 to December 2017, in Delhi NCR, Mumbai, Bangalore, and Kolkata. Each customer account on this food ordering platform has a unique user ID, and we observe the city, subzone, and delivery time of every order made by each account.[3] For each order, we observe the name, unit price, and quantity of each food item. More importantly, we observe payment methods. Customers can pay by cash on delivery, digital wallet, credit card, or internet bank. We consider all non-cash payments as digital payments.

Compared to prior studies on consumer responses to data breaches that use survey data, we perform a more powerful test based on high-frequency transaction-based spending from administrative data directly from the breached firm that have little measurement error.

3. Consumer Responses to the Data Breach

3.1. Change in Payment Modes

We begin our analysis with an unconditional comparison of the two payment modes. The breach was confirmed on the Thursday of week 20 in 2017 (May 18). We drop observations before April 1, 2017, to eliminate the mechanical effect of the November 2016 demonetization on cash availability.[4] We plot the total number of orders paid by digital and cash payments over time in Figure 1. Panel A shows that in the week of the breach, the number of digital payment orders experienced a sharp decline while the number of cash orders increased. This divergence indicates that consumers switched from digital payments to cash payments after the breach. Panel B shows the relative decline in digital payment usage last only temporarily. Digital payment usage gradually recovered, and the gap between digital and cash payments also recovered to the pre-breach level around week 31. We then look into the daily time series seasonally adjusted as

[3] Subzone is a geographic unit in India and is smaller than a city. In our sample, each city has 159.5 subzones on average.

[4] Chodorow-Reich et al. (2020) show that while the demonetization led to an immediate and geographically uneven cash shortage, the cash shortage was alleviated in the subsequent several months. By using a sample period starting in April 2017, more than four months after the demonetization, we by and large remove the impact of demonetization-induced cash shortage.

described in Appendix Section A.1.1. Panel C shows that the gap between the two lines substantially widened on May 18. As time went on, the gap gradually shrank to a negligible level (around August 20, 2017) as shown in Panel D.[5] We examine the number of consumers and the rupee amount of sales for robustness and find similar patterns.

We then rigorously investigate how this exogenous shock changed consumers' payment choices. We construct a balanced panel data set of order characteristics for each payment mode in each week in each subzone and estimate the following difference-in-differences (DID) panel regression:

$$\ln(1 + Y_{ptz}) = \mu_p + \lambda_{tz} + \beta \cdot (Treated_p \times T_t) + \varepsilon_{ptz} \qquad (1)$$

where *p, t,* and *z* represent payment modes (digital or cash), weeks, and subzones, respectively. Y_{ptz} is the number of orders paid by payment mode p in week t in subzone z.[6] We also consider daily frequency in the robustness tests. $Treat_p$ equals 1 if p is digital payment and equals 0 if p is cash payment. T_t equals 1 if week t is equal to or later than week 20 of 2017 in which the breach was announced and equals 0 otherwise. Payment mode fixed effects μ_p control for fixed differences between digital and cash payments and subsume the treatment indicator. Week-subzone fixed effects λ_{tz} control for time trends and allow for heterogeneous time-varying shocks across different subzones. Robust standard errors are clustered at the city level.[7] By this specification, we eliminate the effects of any factors that may affect consumers' willingness to adopt the platform and focus on the difference between the adoption of digital payments and cash payments.

We apply equation (1) to the payment mode-week-subzone panel data set. First, we restrict the sample to the period from week 16 to week 23, i.e., four weeks before and four weeks since week 20. Column 1 of Panel A in Table 1 displays the results. The subzone-level number of digital payment orders declined by 9%, consistent with the short-term decline in Figure 1. We then use observations from week 16 to week 19 (the month before the breach) and from week 28 to week

[5] There was a spike at the end of 2017 in Panel B of Figure 1. There was a similar spike at the end of 2016, which indicates that the spike was a regular seasonal pattern.

[6] We add one to the *Y* to avoid missing value after taking logarithm. We apply this strategy for all regressions that take logarithm to the dependent variables. We also address zero-valued observations using inverse hyperbolic sine transformation and Poisson regression in Appendices B and C.

[7] As the number of cities is low, standard errors clustered at the city level may be downward biased. In our baseline results, we rely on Stata's default multiplicative small-sample correction to the asymptotic variance estimator as well as Stata's default degree of freedom adjustment in the significance testing. We also use the wild cluster bootstrapping tests to correct for the small number of clusters. We also apply the two-way clustering in Appendix D.

31 (the third month after the breach) in Column 2 of Panel A in Table 1. The statistically insignificant coefficient shows that digital payment usages were no longer lower than cash payment adoptions. We find similar patterns in the daily sample in Appendix Table A.1. Compared with cash payments, digital payments decreased in the month after the breach event but recovered in the third month after the breach.

In Panel B of Table 1, we return to the weekly sample and use sales and the price of order (i.e., the value of each payment) as the dependent variable respectively. In the short run (Columns 1 and 2), sales decline by a similar extent to order count, while the price of order does not change. In the third month (Columns 3 and 4), there is little change.

We also analyze using alternative specifications. We exploit a user level sample in Appendix Section A.2.2 and find that consumers were less likely to choose digital payment in the short run after the breach. In the long run, the digital payment adoption recovered. Regarding the identifying assumption, Panel A of Figure 1 shows parallel pre-breach trends in unconditional patterns. We also conduct rigorous regression analysis in Appendix Section A.2.3. In Figure A.3, the coefficients for pre-breach weeks are all statistically and economically indistinguishable from zero, and there is a sharp decline in digital payments immediately after the breach.

3.2. Customer Entries and Quits

Next, we analyze customer entries into the platform. Entrants presumably know about the data breach that happened on the app, given the wide media coverage of the data breach. They did not have their data already on the app. They faced a trade-off between sharing their data to the app and hence being exposed to breaches on the app versus the utility from using the app. Panel A of Figure 2 plots the weekly time series of the number of new customers in the short term. The time series had an upward trend before week 20. The trend was reversed and went down until week 22, which indicates that the data breach discouraged consumers' adoption of this platform. However, the decline only persisted for three weeks, and its impact is small in the long run as shown in Panel B of Figure 2.

We apply our main DID specification to examine entry to the app. For every week and each subzone, we calculate the number of digital payment orders and cash orders by entrants. As can been seen from Appendix Table A.3, the number of digital payment orders by entrants decrease 4.3% per week in the two weeks after the data breach relative to the change for the number of cash

orders by entrants. However, the difference becomes insignificant in the third month after the data breach. The regression results which are obtained with tight fixed effects confirm the unconditional patterns of entry: Digital payments by entrants experiences a dip immediately after the data breach relative to cash payments but rebounds soon.

We also explore how the breach affected the composition of customers entering the platform. In Table 2, we compare the characteristics of the first order made by customers who entered one month before (i.e., between April 18 and May 17) and after the breach (i.e., between May 18 and June 17). *Quantity* is the average number of items in an order. It measures the average size of orders by the user. *Unit Price* is the average unit price of the food items in the orders by the user and acts as a proxy for income level. *Price* is the price of the entire order. *Vegetarian Index* is the ratio between the number of orders without meat and the number of orders with meat or fish. We infer whether an order includes meat or fish by the names of the food items in the order. We use the *Vegetarian Index* to infer the eating habits and investigate composition change of customers. The differences in *Price* and *Vegetarian Index* are statistically insignificant. The differences in *Unit Price* and *Quantity* are statistically significant at the 1% and 10% level respectively, but the magnitude of the differences is negligible: the change in *Unit Price* (*Quantity*) accounts for only 1% (0.8%) of the mean of customers who entered before the breach. Overall, there are little changes in the characteristics of new users' ordering behaviors, suggesting that the composition of new customers likely remains stable after the breach.[8]

We then investigate consumer exits (quits) from the platform. We define quit as a four-week (or longer) inactivity.[9] Specifically, if a user has no orders from week t to week $t + 3$, we consider the user quits in week t. We compute the weekly number of quits and plot the time series in Panel C of Figure 2. There was an upward trend in week 20, but that trend could be due to the momentum of the upward trend that started in week 18. Besides, the trend was persistent only for a couple of weeks, and its magnitude is minor in the long run as displayed in Panel D of Figure 2. The evidence on entries and quits suggests that the data breach had little impact on consumers' adoption of the platform.

[8] We cannot fully reject a change in the customer composition as we do not observe demographic characteristics in our data. Prior research has documented that certain demographic groups, notably female and older people, are more sensitive to data privacy issues than others (Tang, 2019; Prince and Wallsten, 2022).

[9] As a comparison, users who ordered at least once in the fourth month before the breach on average placed one order for every 10.4 days during the following three months.

3.3.Individual Consumer Level Analysis of Quits and Payment Changes

Motivated by the differential reaction between cash and digital payment to the data breach, we explore whether existing consumers change their behaviors after the data breach. Their data on the app were leaked; the possibility that an unauthorized third party may exploit the breached information and bring losses to customers raises concerns for payment security. If they continue to use the app and stick to digital payments after the breach, their activity generates more payment data on the app (such as saved card numbers and reset passwords). For the new data being generated, leakage risks have not been realized but consumer perception of the leakage risks might change because of the breach.

To analyze quits at the individual consumer level, we restrict the analysis to the sample of pre-breach active users, i.e., those who order at least once per week in the 4 weeks before May 18.[10] We construct a cross-sectional sample for these active users with a dummy *Quit* which equals 1 if the consumer quits within four weeks after week 20 and equals 0 otherwise. Since the dependent variable is a dummy, we apply the following logistic regression:

$$\Pr(Quit_i=1) = \Phi(\beta_0 + \beta_1 Digital\_Ratio_i + \sum \gamma_j Control_{ij} + \lambda_c), \quad (2)$$

where $\Phi(x) = \exp(x)/(1+\exp(x))$.

The key independent variable is *Digital_Ratio*, the ratio between the number of digital payments and the total number of payments from April 1 to May 17 of 2017. The variable measures the extent to which the consumer relies on digital payments. λ_c is city fixed effects. We include the following control variables: *Quantity, Unit price, Vegetarian Index, Account age,* and *Frequency. Account age* is the number of weeks between the first order by the user and week 18 of 2017. This variable is a proxy for how long the users had adopted the platform for. *Frequency* is the average number of orders per week by the user (during the period from his first order to May 18, 2017. The other three variables are constructed analogously as before.

In Table 3, Panel A shows the summary statistics and Column 1 of Panel B displays the estimation results. The average marginal effects are -0.103, which indicates a negative correlation between the odds of quits and *Digital_Ratio*. The results change little qualitatively when we include all controls in Column 2.

[10] Note that May 18 (the date of breach announcement) was the Thursday of week 20. Even if a consumer wanted to quit immediately after the breach, she may still have had orders in week 20 since we use weekly sample here. Hence, the earliest week for quits is week 21.

We conduct an additional test to show that digital payment users switched to cash payments:

$$\Delta Digital\_Ratio = \beta_0 + \beta_1 Digital\_Ratio_i + \sum \gamma_j Control_{ij} + \lambda_c + \varepsilon_i, \quad (3)$$

We restrict the sample to "active" users identical to those in equation (2) and drop users who quit in week 21 (who hence have no payment records after week 21). Then we compute the *ΔDigital_Ratio* which equals the *Digital_Ratio* during the period from week 21 to week 24 (i.e., the month after the breach) minus the *Digital_Ratio* used in equation (2) (i.e., the ratio before the breach). The independent variables are similar to those in equation (2). Columns 3 and 4 of Panel B in Table 3 display the results. The decline in the digital payment ratio was greater for consumers with higher digital payment ratios. In other words, they switched to cash to a larger extent. To eliminate the possibility that the greater decline is due to higher baseline of consumers with higher digital payment ratios, we replace the dependent variable with the natural logarithm of the digital payment count to measure growth rates. Columns 5 and 6 in Panel B of Table 3 report the estimates which indicate similar results.[11]

Columns 3 to 6 of Panel B in Table 3 indicate that digital payment users who prefer convenience more were reluctant to leave the platform, while cash users who prefer convenience less were more likely to quit the platform. Instead of immediately quitting, digital payment users sacrifice the convenience of digital payments to secure payment security but still retain the convenience of the takeout ordering platform.

4. Discussions of Potential Mechanisms for the Observed Consumer Responses

Compared to survey-based studies, our approach of analyzing consumer activity does not rely on participant self-report and recall and therefore is not subject to being incomplete due to limits of memory, under-reported due to forgetting past experiences, or over-reported due to being influenced by media coverage. A limitation is that we are unable to directly measure consumer perceptions from actual behaviors. In this section, we discuss potential mechanisms through which the data breach affected consumer behaviors.

The data breach led to an immediate decrease in digital payments relative to cash payment despite that payment information was not leaked. One possibility is that initially the app users thought hackers could capture payment information, they switched to using cash payments until

[11] Note the value of *Digital_Ratio* is between 0 and 1, and the coefficient measure the marginal effect when *Digital_Ratio* increases from 0 to 1, and hence the big coefficients in Columns 5 and 6 are not surprising.

they realized the breach did not jeopardize their payment security.[12] Given that the initial discovery of the breach was limited to observers of the dark web marketplace and the company confirmed the data breach on the next day, most users learned about the breach from the platform. We believe that users were aware of the platform's claim that payment information was not leaked. It is possible that despite the platform's claim, users might be wary of payment security. Such concerns for payment security drive the initial response of digital payments.

The recovery of consumer demand to pre-breach levels by the third month after the breach shows that the impact of the data breach is transient as opposed to permanent. We consider several explanations for this transient effect. Firstly, we can assess the effect through the lens of a trade-off between convenience and payment security. When personal information is leaked in a data breach, consumers' decisions of changing their usage patterns depend on how much utility they gain from the convenience of using the app relative to the disutility associated with the breach. Observed behaviors allow us to indirectly infer consumer preferences: While the concerns for payment security triggered by the privacy leakage lead to a switch away from digital payments in the short term, the perceived benefits of convenience outweigh the costs in the longer term; as result, usages return to pre-breach levels. Admittedly, the costs can evolve over time due to either changes in the breach severity or consumer perception changes. Our analysis does not disentangle these different mechanisms. We leave this interesting area for future research.

Secondly, we consider the possibility that the breached firm engages in safety measures ex post. We searched news outlets and various Internet sources to track the firm's response to the breach. The firm contacted the hacker to delete the breached data and introduced monetary incentives for reporting security flaws in its existing bug bounty program. Such measures signal that the firm is increasing cybersecurity investments to reduce future breach risks and can potentially mitigate consumer concerns for future data breaches. In general, it is difficult for the increased cybersecurity investments to neutralize all future data breaches due to the enormity of potential system vulnerabilities and as a result, repeated data breaches are common. In Appendix A.3.2, we further discuss repeated data breaches and the firm's mitigation measures and acknowledge that the measures by the breached firm may partially mitigate consumers' concerns

[12] It is also possible that the app users may under-estimate how damaging the data breach was and under-react initially. Under this alternative scenario, they should know better over time and correct their initial under-reaction. If this is the case, we would have observed persistent or even increasing reaction to the data breach. We do not see this in the data.

about future data breaches. In addition, we find no evidence that the food delivery platform tried to retain its users by promotion using Google search index in Appendix Section A.2.6.

We also consider the role of inattention, that is, only a small fraction of the app users were aware of the data breach. At its face value, this seems unlikely given the large and significant drop in electronic payment immediately following the breach, and that the company had reset the passwords for all affected users and logged them out of the app and website. If, for some reasons, only a small fraction of the app users learned about the data breach initially, more app users should become aware of the data breach over time, leading to a persistent or even increasing reaction to the data breach. We do not see this in the data. Therefore, the only plausible form of inattention to explain our findings is that app users become inattentive over time (i.e., they forget). This explanation of limited memory is consistent with the trade-off interpretation: the affected consumers do not view the data breach as a big deal, otherwise they would not forget. We also exploit the geographic variation in attention as measured by Google search volume to test the role of inattention in Appendix Section A.3.1 and find no support for the inattention explanation.

Finally, the industrial organization structure of the food delivery industry may affect our findings. If there was only one major and reliable platform in a city, it could be mechanical that the consumers cannot migrate to other options. In this case, the limited options rather than the choice for convenience drive our results. This is unlikely to be the case as there were at least three major food delivery platforms, including the breached platform, in each of the four cities included in our sample. Therefore, consumers can migrate to another platform if they wish to.

5. Additional Analyses and Discussions

Other data breaches. We also analyze the consumer response to two other unexpected data breaches using administrative data from the breached firms. The two breaches involved a leading online grocery store and a leading national bank in India. They stem from different causes, but similar to the food delivery platform's breach as well as the majority of other data breaches, they involve what is typically considered non-sensitive data. In these two data breaches, we find weaker responses (Appendix Sections A.4 and A.5). We also conduct additional robustness checks by restricting the samples to the two overlapped cities Mumbai and Bangalore in Appendix E and find similar results. By studying multiple unexpected data breaches across different industries, we ensure that our results are not driven by app- or industry-specific idiosyncratic characteristics.

Statistical inference. In our baseline results, we rely on Stata's default small-sample corrections to mitigate the potential downward bias of clustered standard errors when the number of clusters is low. We also use the wild cluster bootstrapping tests and show robustness of our statistical inference to computing two-way clustered standard errors in Appendix D. The largely null results raise another concern of whether the effect is close to 0 or the data does not have enough statistical power. To address this concern, we implement the "serial-correlation-robust" (SCR) power calculations (Burlig, Preonas, and Woerman, 2020), which allow for arbitrary (non-constant) serial correlation in panel data applications, for our regression analyses in Appendix Section A.6.1. We show that our null results are not driven by insufficient statistical power. In our setting, a few factors boost the statistical power. First, the data were collected directly from the apps' transaction records and have very little (if any) measurement error. Second, we use weekly or daily data in our analysis, which provided high-frequency observations to detect the effect of data breaches.

External validity. We draw our findings from multiple data breaches in India. For the following reasons, we believe that our findings can be applicable to other data breaches and other countries. First, similar to the data breaches that we analyze, the majority of data breaches do not involve leakage of sensitive personal information. Second, existing studies in household finance (e.g., D'Acunto, Prabhala, and Rossi, 2019) document that Indian consumers behave similarly to those in other countries. Third, a number of survey and experimental studies conducted in the U.S. and other countries show a paradox that consumers' self-reported preferences for sharing personal data are inconsistent with their actual behaviors. By analyzing consumer transactions, our study complements these survey and experimental studies to provide additional empirical estimates for consumers' personal data sharing preferences and how the preferences affect their data-sharing choices. In Appendix Section A.6.2, we provide additional details and discussions for external validity.

Implications for data protection policies. In recent years, growing concerns for consumer privacy have prompted the enactment of data protection laws such as EU's General Data Protection Regulation (GDPR). An important objective of these laws is to restrict collection of personal data, which can be misused for unwarranted reasons (e.g., discrimination, exploitation of consumer vulnerability) or leaked in cybersecurity breaches. These undesirable costs need to be weighed against the improved access to products and services and the substantial convenience

enabled by big data technologies. In this study, we focus on the setting of data breaches, a risk factor from which data protection laws aim at protecting consumers, to analyze this trade-off. By inferring consumer preferences from observed behaviors in unexpected data breaches, we empirically demonstrate that for consumers, the benefits of convenience outweigh the cost of concerns for payment security. Our findings are related to Liu, Sockin, and Xiong (2021)'s theoretical framework that models consumer preference for personal data sharing as a mechanism for concealing behavioral vulnerabilities. In their framework, restricting data sharing is more beneficial to consumers for temptation goods such as gambling or video games; for normal consumption goods such as food delivery, grocery, and banking service that we study in this paper, convenience dominates the concern for personal information security. In Appendix Section A.6.2, we provide additional discussions for the policy implications.

6. Conclusion

This paper investigates how consumers respond to unexpected data breaches of their personal information. We find that after an unexpected data breach, customers of a food delivery platform switched from digital payments to cash in the short run. In the long run, usage of digital payments recovered. Customers' entry and exit intensities changed little following the data breach. Existing users who primarily used digital payments are less likely to quit the platform following the data breach, compared with those who primarily used cash. Additional analyses on bank and online grocery data breaches uncover even weaker effects. Overall, our analyses show that data breaches are indeed recognized by affected users in the short term. In the long run, usage returns to pre-breach levels.

Our analyses utilize actual transaction data to contribute to our understanding of how data breaches affect consumer behaviors. Our findings imply that the perceived benefits of convenience outweigh the costs of payment security risks. The transient impact of data breaches on consumer behaviors is conceptually applicable to other data breaches of non-sensitive data. Consumers may respond more strongly to data breaches that compromise financial or personally identifying data.

References

Argente, David, Chang-Tai Hsieh, and Munseob Lee. 2022. "The cost of privacy: welfare effects of the disclosure of Covid-19 cases." *Review of Economics and Statistics* no. 104 (1):176-186.

Burlig, Fiona, Louis Preonas, and Matt Woerman. 2020. "Panel data and experimental design." *Journal of Development Economics* no. 144:102458. doi: 10.1016/j.jdeveco.2020.102458.

Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra, and Abhinav Narayanan. 2020. "Cash and the economy: Evidence from India's demonetization." The Quarterly Journal of Economics no. 135 (1):57-103.

Cong, Lin William, Danxia Xie, and Longtian Zhang. 2021. "Knowledge Accumulation, Privacy, and Growth in a Data Economy." *Management Science* no. 67:6480-6492. doi: 10.1287/mnsc.2021.3986.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G Rossi. 2019. "The promises and pitfalls of robo-advising." *The Review of Financial Studies* no. 32 (5):1983-2020.

Greene, Claire, and Joanna Stavins. 2017. "Did the Target data breach change consumer assessments of payment card security?" *Journal of Payments Strategy and Systems* no. 11 (2):121-133.

Gu, Jie, Yunjie Calvin Xu, Heng Xu, Cheng Zhang, and Hong Ling. 2017. "Privacy concerns for mobile app download: An elaboration likelihood model perspective." *Decision Support Systems* no. 94:19-28.

Janakiraman, Ramkumar, Joon Ho Lim, and Rishika Rishika. 2018. "The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer." *Journal of Marketing* no. 82 (2):85-105.

Jones, Charles I, and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." *American Economic Review* no. 110 (9):2819-58.

Kwon, Juhee, and M Eric Johnson. 2015. The market effect of healthcare security: Do patients care about data breaches? Paper read at WEIS.

Liu, John Zhuang, Michael Sockin, and Wei Xiong. 2021. Data Privacy and Consumer Vulnerability.

Mikhed, Vyacheslav, and Michael Vogan. 2018. "How data breaches affect consumer credit." *Journal of Banking Finance* no. 88:192-207.

Pentina, Iryna, Lixuan Zhang, Hatem Bata, and Ying Chen. 2016. "Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison." *Computers in Human Behavior* no. 65:409-419.

Prince, Jeffrey T, and Scott Wallsten. 2022. "How much is privacy worth around the world and across platforms?" *Journal of Economics & Management Strategy* no. 31 (4):841-861.

Rysman, Marc, and Scott Schuh. 2017. "New innovations in payments." *Innovation Policy and the Economy* no. 17 (1):27-48.

Schuh, Scott, and Joanna Stavins. 2016. "How do speed and security influence consumers'payment behavior?" *Contemporary Economic Policy* no. 34 (4):595-613.

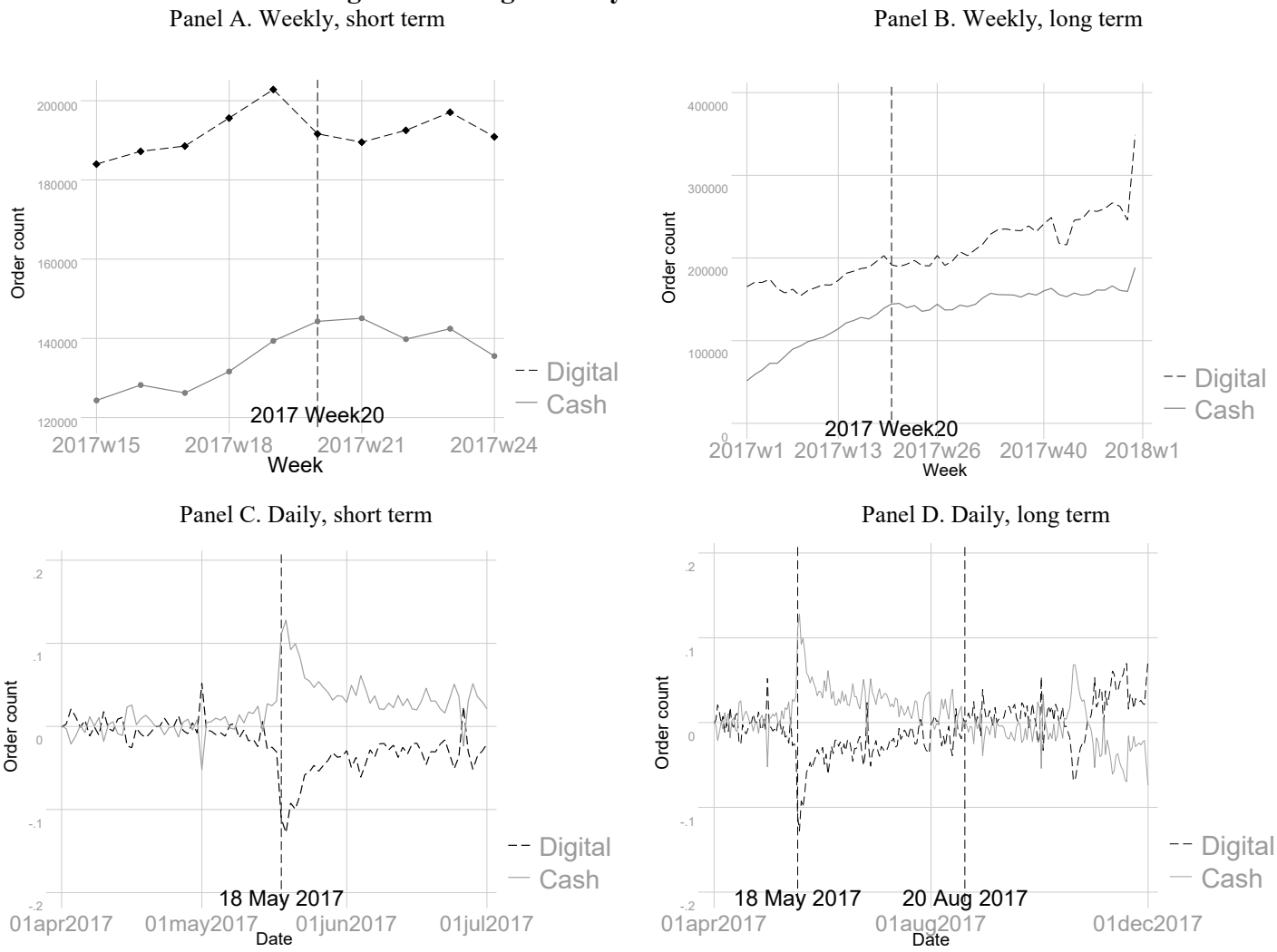
Tang, Huan. 2019. The Value of Privacy: Evidence from Online Borrowers.

Teo, Aik-Chuan, Garry Wei-Han Tan, Keng-Boon Ooi, Teck-Soon Hew, and King-Tak Yew. 2015. "The effects of convenience and speed in m-payment." *Industrial Management & Data Systems*.

U.S. Government Accountability Office. 2007. Personal Information: Data breaches are frequent, but evidence of resulting identity theft is limited; however, the full extent is unknown.

Wottrich, Verena M, Eva A van Reijmersdal, and Edith G Smit. 2018. "The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns." *Decision support systems* no. 106:44-52.

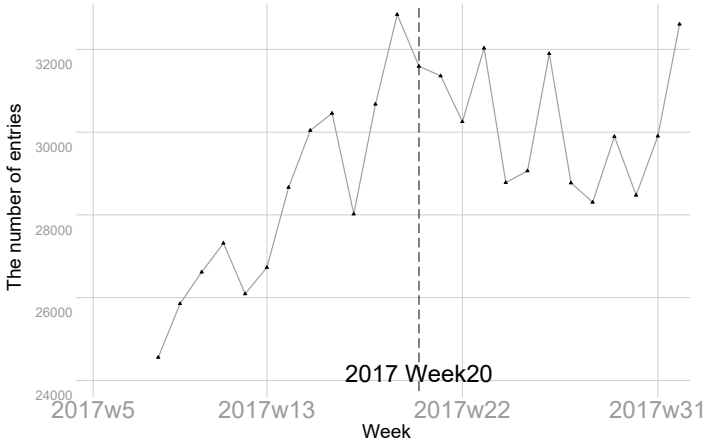
Figure 1 Changes in Payment Modes – Order Count



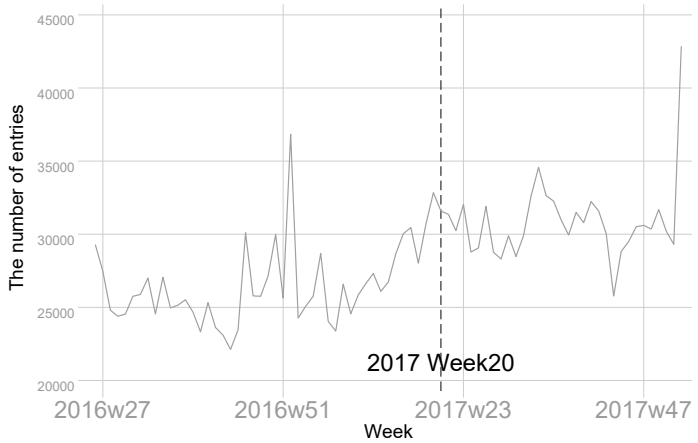
The figure shows the time series of the number of orders paid by digital and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample. Panels C and D are for the seasonally adjusted daily sample mentioned in Section 3.1.

Figure 2 Consumer Entries and Quits

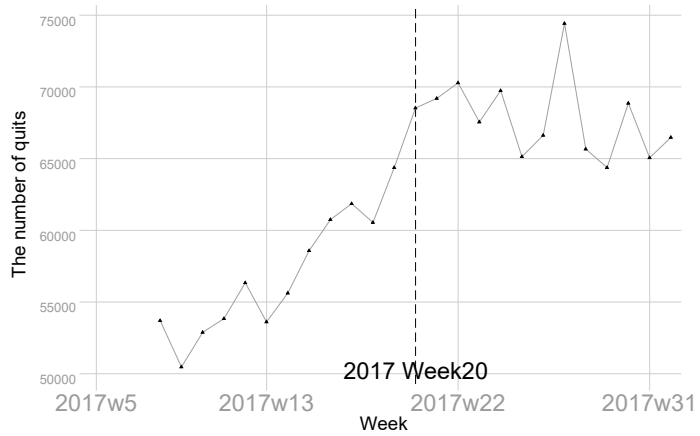
Panel A. Consumer Entries, short term



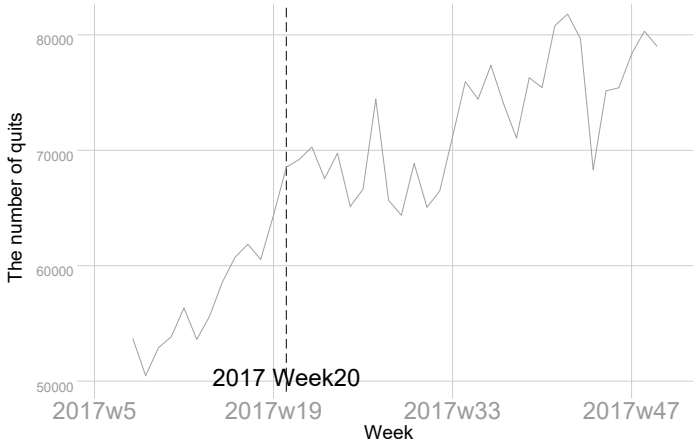
Panel B. Consumer Entries, long term



Panel C. Consumer Quits, short term



Panel D. Consumer Quits, long term



The figure shows the time series of consumer entries and quits. Panels A and B are for entries that are defined as the first appearance of the user in the sample. Panels C and D are for quits. The week of a quit is defined as the first week of four-week inactivity.

Table 1 Changes in Payment Modes
Panel A Weekly

Dep. var.: Ln(1+order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.090***	0.019
	(-6.32)	(0.75)
Payment Mode FE (Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	10,224	10,224
Observations with DV=0	944	730
R^2	0.976	0.975
Mean DV	4.081	4.156
P value for wild-t	0.058	0.529

Panel B Alternative Dependent Variables

Dep. var.:	(1) Ln(1+sales)	(2) Ln(1+order price)	(3) Ln(1+sales)	(4) Ln(1+order price)
	Post: the first month		Post: the third month	
Post × Digital	-0.101**	-0.004	0.092	0.020
	(-3.51)	(-0.31)	(1.46)	(1.37)
Payment Mode FE(Digital dummy)	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes
Observations	10,224	9,090	10,096	9,332
Observations with DV=0	944	0	730	0
R^2	0.952	0.817	0.951	0.815
Mean DV	9.455	5.977	9.638	5.960
P value for wild-t	0.064	0.850	0.056	0.166

Subzone-payment mode-weekly (daily) level regressions estimating the response of payment modes to the data breach of the food delivery platform. In Panel A, we compare the month before the breach with the first and the third month after the breach respectively, and the dependent variable is the natural logarithm of the number of orders. In Panel B, we use the two sample periods respectively and replace the dependent variable with sales and the price of orders, respectively. The mean dependent variable, $\ln(1+y)$, is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where G is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 2 Data Breach and Customer Entries

Variable	Before	After	Diff/P-value
Price	1,507.613	1,489.669	-17.944
			[0.241]
Unit Price	181.840	183.839	1.999***
			[0.000004]
Quantity	3.165	3.141	-0.024*
			[0.069]
Vegetarian Index	0.360	0.357	-0.003*
			[0.099]
Observations	130,738	135,600	266,338

This table shows the t-tests comparing the means of the characteristics of the first order by the customers who entered one month before and after the data breach.

Table 3 Quits and Payment Switches

Panel A Summary Statistics

Variable	Obs	Mean	Std. Dev.
Quit	28,930	0.18	0.38
Digital_ratio	28,930	0.63	0.30
Quantity	28,930	2.71	1.18
Account Age	28,930	39.36	17.56
Frequency	28,930	1.60	1.19
Unit price	28,930	164.63	49.89
Vegetarian Index	28,930	0.18	0.15
ΔDigital_Ratio	27,117	0.01	0.26

Panel B Regression Results

Dep. var.	Quit LOGIT Coef.		ΔDigital_Ratio OLS Coef.		ΔLn(1+Digital payments) OLS Coef.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digtal_ratio	-0.711***	-0.484***	-0.089***	-0.097***	-1.163***	-1.002***
	(-14.16)	(-18.46)	(-6.10)	(-6.65)	(-50.22)	(-40.17)
Quantity		-0.014		0.002		0.003
		(-0.93)		(1.53)		(1.53)
Account Age		-0.028***		0.002***		-0.027***
		(-18.04)		(11.34)		(-55.24)
Frequency		-0.378***		0.005*		-0.182***
		(-39.99)		(2.66)		(-11.13)
Unit price		-0.003***		0.000***		0.000*
		(-3.21)		(6.52)		(2.42)
Vegetarian index		-0.015		0.006		0.076
		(-0.30)		(0.58)		(1.24)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,930	28,930	27,117	27,117	27,117	27,117
R^2			0.013	0.024	0.137	0.372
Pseudo R^2	0.014	0.063				
Mean DV	0.179	0.179	0.014	0.014	-1.641	-1.641
P-value of wild-t	0.093	0.027	0.086	0.084	0.011	0.019
Marginal Effects of	-0.103***	-0.067***				
Digital_Ratio	(-14.38)	(-19.53)				

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. Panel A displays the descriptive statistics of the sample, and Panel B displays the estimates. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the natural logarithm of the digital payment. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. The mean dependent variable is reported at the bottom to assess marginal effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Internet Appendix for
Transient Customer Response to Data Breaches of Their Information

Sumit Agarwal, Pulak Ghosh, Tianyue Ruan, and Yunqi Zhang*

This Internet Appendix contains additional material, tables, and figures referenced in the main text.

A. Appendix A

A.1 Methodology referenced in the main text

A.1.1 Seasonality Adjustment for the Daily Sample

This section introduces how to construct seasonally adjusted daily time series. We first construct a daily-payment mode level panel dataset. We run the following regression to eliminate seasonality:

$$Ln(Y_{pt}) = \beta_0 + \lambda_{pd} + \mu_t + \varepsilon_{pt}, \quad \text{(A.1)}$$

where p, t, and d represent payment mode, date, and day of week, respectively. Y_{pt} is the value of the dependent variables we are interested in. λ_{pd} is the payment mode - day of week fixed effects.[1] λ_{pd} controls for the within-week cyclicality that can potentially differ between digital and cash payments. μ_t controls for the date fixed effects. We drop observations before April 1, 2017, to eliminate the effects of the demonetization in India.[2] We then use the residuals ε_{pt} as the time series of daily order count, one for digital payments (ε_{pt} where p is digital payment), the other for cash payments ($\boldsymbol{\varepsilon_{pt}}$ where p is cash payment).

A.1.2 Demonetization

The Demonetization in 2016 had a huge impact on payment choices. It began on November 9, 2016, in which the old version of the 500- and 1,000-rupee notes could not be used in transactions and were replaced by new 500- and 2,000-rupee notes. This demonetization led to a cash shortage and forced people to turn to digital payments (Agarwal et al., 2019; Chodorow-Reich et al., 2020). As time goes on, the cash shortage was alleviated. In Figure A.1, we use the full sample to construct the two time series of cash and digital payments and plot them. Panel A of Figure A.1 shows substantial changes in the two payment modes around November 9, 2016. Then, as the cash shortage was alleviated, digital (cash) payments declined (rose), and the gap between the two payment modes became stable around April 1, 2017. There was a small gap between the two payment modes before the breach event (Panel B of Figure A.1). To eliminate the gap and illustrate the dynamic change in the difference between the two payment modes more clearly, we drop observations before April 1, 2017. However, Panel B of Figure A.1 shows the conclusion does not change even if we use the full sample.

* Sumit Agarwal (ushakri@yahoo.com) and Tianyue Ruan (tianyue.ruan@nus.edu.sg) are from the National University of Singapore, Pulak Ghosh is from the Indian Institute of Management Bangalore (pulakghoshc@gmail.com), and Yunqi Zhang is from Nankai University (zhangyunqi@u.nus.edu).

[1] The samples for most figures in the paper do not have zero observations for the dependent variable since we aggregate the variable at the time-payment mode level, The exceptions are Figure 1 (Panels C and D) and Figure A.1 for which we add one to the Y to avoid missing value after taking logarithm.

[2] We discuss how we handle demonetization in Section A.1.2.

Figure A.2 displays the weekly time series of the number of orders for cash and digital payments respectively using the full sample. There were spikes at the end of 2016 and 2017.

A.2 Additional Tests

A.2.1 Daily-level Analysis: Change in Payment Modes for The Food Delivery Platform

We use alternative specifications to verify the findings in Section 3.1. We first use the subzone-daily level sample on the order placing and payments modes. We restrict the sample to users who ordered at least once during the month right before the breach. Then we run the following two regressions:

$$\ln(1+Y_{zd}) = \beta_0 + Post_d \times Digital_p + \mu_z + \varepsilon_{zd}, \quad \text{(A.2)}$$

where *z*, *d*, and *p* represent user, date, and payment mode, respectively. Y_{zd} is the number of orders placed by the user u during day *d*. We add one to Y_{zd} before taking logarithm since there are many observations with zero values for this variable. $Post_d$ is a dummy which equals 0 if the week is before the breach and 0 otherwise. $Digital_p$ is a dummy which equals 1 if the payment is digital. μ_z controls for subzone fixed effects. Robust standard errors are clustered at the city level for both regressions. The short-term sample spans from 1 April 2017 to 4 weeks after the week of breach. The long-term sample includes the period from week 16 to week 19 (the month before the breach) and the period from week 28 to week 31 (the third month after the breach), which is similar to the sample period of Panel B of Table 1. Table A.1 displays the results. Particularly, Column (1) in shows that there was a 6.0% decline in digital payment usage in the short run. However, in the long run, digital payment usage reversed (Column 2), which is consistent the results in Section 3.1.

A.2.2 User-level Analysis: Change in Payment Modes for The Food Delivery Platform

We also construct a balanced user-week level sample of food delivery platform on the order placing and payments modes. We restrict the sample to users who ordered at least once during the month right before the breach. Then we run the following two regressions:

$$\ln(1+Y_{uw}) = \beta_0 + Post_w \times Digital_p + \mu_u + \varepsilon_{uw}, \quad \text{(A.3a)}$$

$$Digit\_Share_{uw} = \beta_0 + Post_w + \mu_u + \varepsilon_{uw}, \quad \text{(A.3b)}$$

where *u*, *w*, and *p* represent user, week, and payment mode, respectively. Y_{uw} is the number of orders placed by the user u during week *w*. We add one to Y_{uw} before taking logarithm since there are many observations with zero values for this variable. $Post_w$ is a dummy which equals 0 if the week is before the breach and 0 otherwise. $Digital_p$ is a dummy which equals 1 if the payment is digital. μ_u controls for user fixed effects. $Digit\_Share_{uw}$ is the ratio digital payments to the total number of payments for user *u* in week *w*. Note $Digit\_Share_{uw}$ can only be calculated when Y_{uw} is non-zero, and hence regression A.3b is only for observations with for observations with positive Y_{uw}. Robust standard errors are clustered at the user level for both regressions. Note the second regression is conditional on placing orders. We use two sample period respectively. The short-term sample spans from 1 April 2017 to 4 weeks after the week of breach. The long-term sample includes the period from week 16 to week 19 (the month before the breach) and the period from week 28 to week 31 (the third month after the breach), which is similar to the sample period of Panel B of Table 1. Table A.2 displays the results. Particularly, Column (1) in Panel A shows that

there was a 1.2% decline in digital payment usage in the short run. However, in the long run, digital payment usage reversed (Column 2 of Panel A), which is consistent the results in Section 3.1. Panel B shows a similar reversal pattern.

A.2.3 Event Study for The Food Delivery Platform

We then validate the parallel trends assumption by the following event study regression model in the payment mode-week-user panel data set:

$$\ln(1 + Y_{ptu}) = \mu_p + \lambda_{tu} + \sum_{t=-3}^{11} \beta_t \cdot (Treated_p \times \mathrm{I}_t) + \varepsilon_{ptu} \qquad \text{(A.4)}$$

where *p*, *t*, and *u* represent payment modes (digital or cash), weeks, and users, respectively. Y_{ptu} is user *u*'s number of orders paid by payment mode *p* in week *t*. $Treat_p$ equals 1 if *p* is digital payment and equals 0 if *p* is cash payment. I_t is an indicator variable for each of the weeks around week 20, the week of data breach. The sample period includes the four weeks before the breach and the twelve weeks afterwards, from week 16 to week 31. The fourth week prior to the data breach (week 16) constitutes the omitted baseline group. The coefficient β_0 measures the response of digital payments relative to cash payments in the week of the data breach, week 20. $\beta_1, \cdots, \beta_{11}$ track the response of digital payments relative to cash payments one week, two weeks, ⋯, and eleven weeks after the breach week, respectively. Similarly, $\beta_{-3}, \cdots, \beta_{-1}$ capture the difference of trends between digital payments and cash payments in each of the three weeks before the data breach.

Figure A.3 presents the entire path of estimated coefficients β_t along with the associated 95% confidence intervals for the log of the number of orders plus one. The coefficients for pre-data breach weeks are all statistically insignificant. More importantly, we can see a sharp decline in the week of breach, hence the decline is unlikely due to a pre-existing trend. The use of weekly frequency implies that the immediate response to the data breach is split between β_0 and β_1, as May 18 (the date of breach announcement) is the Thursday of week 20. Even if a consumer decreased digital spending immediately after the breach, she may still have digital spending from the days before the breach as recorded as week 20 spending. Since the second week after the breach, use of digital payments has gradually returned to the pre-breach level. The event study estimates corroborate the unconditional patterns in Panel A of Figure 1 to establish the validity of the parallel trends assumption.

A.2.4 Ex ante versus ex post and Extensive Margin versus Intensive Margin

In this section, we decompose the data into existing users and entrants to further validate the trade-off argument.

New entrants presumably know about the data breach that happened on the app, given the wide media coverage of the data breach. These new entrants do not have their data already on the app. They face a clear trade-off between exposing their data to the app and hence exposed to breaches on the app versus the convenience of using the app. Thus, this group provides a clear setting to test the trade-off hypothesis.

For every week and each subzone, we calculate the number of digital payment orders and cash orders by entrants. The entrants are defined as the observations for the first week each use appear in the platform. We first aggregate the weekly entrants' order count for digital and cash payment respectively and plot the two time series in Figure A.4. In week 21 (the week following the data breach), the orders by cash payment increased while the orders with digital payments declined.

We apply our main diff-in-diff specification (equation (2) in the main text) to examine entry to the app. The dependent variable is replaced with the number of digital payment orders and cash orders by entrants. The results are displayed in of Table A.3.

The number of digital payment orders by entrants decrease 4.3% per week (Column 2) in the two weeks after the data breach (compared to the two weeks before the data breach) relative to the change for the number of cash orders by entrants. However, the difference becomes insignificant in the third month (Column 3) after the data breach.[3] The regression results which are obtained with tight fixed effects confirm the unconditional patterns in Figure 2: Entry to the app experiences a dip immediately after the data breach, but rebounds fairly soon.

Above are evidence for extensive margin. We also investigate the intensive margin by focusing on existing users. In fact, we have analyzed existing users at the user-week level in Columns 1 and 2 in Panel A of Table A.2 in which we restrict the sample to users who ordered at least once during the month right before the breach. In Columns 4 to 6 of Table A.3, we use subzone-week level sample of existing users to conduct the diff-in-diff analyses (equation (2) in the main text). The results suggest that the intensive margin also exists: digital payments initially declined and then recovered.

A.2.5 Consumer Count and Sales

In Figure A.5 and Figure A.6, we using the method in Section 3.1 to construct both weekly and daily time series for consumer count and sales, respectively. More specifically, we calculate the number of users who used the cash (digital) payments for at least once in week *t*. Therefore, we can obtain a weekly time series for cash (digital) users. Similarly, we can obtain the daily time series for cash and digital users and seasonally adjust them using the method in Section A.1.1. For sales, we observe the unit price and quantity of each food item in an order, and hence we can calculate the total price of the order. Subsequently, we calculate the weekly and the (adjusted) daily time series of the rupee volume of orders paid by cash/digital payments. The pattern is similar to the one introduced in Section 3.1. The gap between the two payment modes was substantially widened on the day of the data breach and gradually shrank to a negligible level in the long run.

A.2.6 Promotion

We search promotion information on Google and focus on the period from May 17, 2017, to August 20, 2017 (three months after the event). Google has a search syntax where "A"+"B" would generate web pages containing both word A and word B. We set the region to India and tried several combinations involving the name of the platform: "the platform name"+"promotion"+"breach", "the platform name"+"discount"+"breach", and "the platform name"+"coupon"+"breach". We enter each link in the search result and read the contents of the webpage. We find no news about promotions associated with the data breach.

It is possible that the platform increased promotion but did not market it as related to the breach. We manually collect and plot a time series for the weekly number of Google search results for "the platform name"+"coupon", and try to see whether the available coupons increased after the event (Figure A.7). Note the event was in week 20 of 2017. Before week 20, there was an upward

[3] Similar to the baseline regression, we also use one month before and after the breach as the sample period for entrants in Column 2. The coefficient is only significant at the 10% level, which implies entrants shift back to cash more quickly.

trend that was reversed in week 20. Then the number seasonally went up and down. In other words, there is no clear pattern of increases in promotions.

A.3. Details of Some Potential Mechanisms

This section provides additional details for several potential mechanisms mention in the main text.

A.3.1 Inattention

One may argue that the weak response is due to users' inattention to the breach, that is, only a small fraction of the app users are aware of the data breach. First, this seems unlikely given the large and significant drop in terms of electronic payment immediately following the breach that we find, and that the company claimed that it had reset the passwords for all affected users and logged them out of the app and website.

Second, if, for some reasons, only a small fraction of the app users learned about the data breach initially, more app users should become aware of the data breach over time. If this is the case and concerns for payment security are severe, we would have observed persistent or even increasing reaction to the data breach. We do not see this in the data. Third, perhaps app users become inattentive over time (i.e., they forget), but then this is consistent with our interpretation of trade-off (they do not think this is such a big deal, otherwise they would not forget).

We also exploit the geographic variation in attention to test the inattention explanation. We use a Google search syntax "A"+"B" similar to the one in Section A.5.1. We set the region to each city, respectively, and manually collect the weekly search index (the number of Google search results) for "the platform name" + "breach".[4] We then use a triple difference involving the city-weekly search index in the main regression as follows:

$$\text{Ln}(1+Y_{ptz}) = \beta_0 + \beta_1 Treat_p \times T_t \times Google\ search_{ct} + \beta_2 Treat_p \times T_t + \beta_3 Treat_p \times Google\ search_{ct} + \beta_4 Treat_p \times Google\ search_{ct} + \beta_5 Treat_p + \beta_6 Google\ search_{ct} + \lambda_{tz} + \varepsilon_{ptz}, \quad \text{(A.5)}$$

where c denotes cities. Other variables and notations are the same to equation (2) in the main text. Table A.4 presents the results. The coefficient of the triple interaction term is statistically and economically insignificant, which does not support the inattention explanation.

A.3.2 Repeated Data Breaches and Firm Mitigation Measures

In this section, we discuss how firm's response can mitigate the concern of future breaches. In general, while cybersecurity investments cannot technically avoid all future data breaches, they may partially mitigate consumers' concerns about future data breaches.

Cybersecurity investments cannot neutralize all future data breaches due to the enormity of potential system vulnerabilities. No two data breaches are alike. The four data breaches we analyze in our paper stemmed from different causes. We further review the causes of data breaches based on data published by the Privacy Rights Clearinghouse, an organization that tracks data breaches investigated by U.S. state Attorneys General and the U.S. Department of Health and Human Services. In the PRC chronology of data breaches (https://privacyrights.org/data-breaches), there are 8,937 data breaches that took place during the 14-year period from 2005 to 2018. Table A.5

[4] Google does not provide data at the subzone level.

tabulates the causes of these 8,937 data breaches as classified by the PRC. 91.12% of the breaches are classified into one of seven causes: hacking/malware; unintentional disclosure; loss or theft of physical materials; loss or theft of portable devices; insider leakage; loss or theft of stationary computer; and fraudulent card skimming. For the other 8.88% of the breaches, there was not enough information about breach to know how exactly the information was exposed. Considering that the PRC sources the information from official investigations that typically involve forensic analysis, the non-trivial portion of data breaches with unknown causes highlights the challenges in diagnosing data breaches.

In sum, data breaches are triggered by vastly different causes and therefore are difficult to eliminate ex ante. Hackers are perceived to be always ahead. In 2016, 65% of cybersecurity officials from U.S. Defense Department, civilian and intelligence agencies said they disagreed with the idea that the federal government can detect cyberattacks while they're happening. In addition, 59% said that their "agency struggles to understand how cyber attackers could potentially breach their systems".

As a result, despite the breached companies' potentially higher cybersecurity investments, repeated data breaches are common. In the PRC data, 940 companies (12.4%) experienced repeated data breaches between 2005 and 2018. As a comparison, 6,633 companies (87.6%) experienced only one data breach during this period. In a November 2021 survey administered among small, medium, and large-size U.S. companies in consumer-facing industries (retail, food/beverage, and hospitality), 31% of the 300 surveyed companies experienced a data breach in the past; among companies that had suffered a breach, 89% had experienced multiple breaches in a single year. Repeated data breaches can affect even the largest and most well-known companies that are least likely to be financially constrained for cybersecurity investments. For instance, Yahoo! experienced data breaches in 2013, 2014, and another data breach that went undetected until 2016. T-Mobile reported four data breaches within the three-year period from 2018 to 2020. Sony experienced a data breach in the PlayStation networks in April 2011 and then another later breach in its film subsidiary, Sony Pictures, in November 2014.

Furthermore, in repeated data breaches, the severity of the later breaches may not decline either. The last breach in the streak of T-Mobile data breaches involved data on calls including numbers called, frequency, duration, and timing of the calls, or Customer Proprietary Network Information (CPNI), considered by the FCC to be some of the most sensitive personal information that wireless service providers collect about their customers and accounts for actual user activity.

On the hand, depending on the breached firm's responses and measures, the occurrence of one data breach may influence consumers to perceive future data breaches differently. To shed light on this, we searched the breached firm's name in India news outlets and other Internet sources for the three-month period following the revelation of the breach and analyzed all the relevant statements, reports, and web records. We find the following mitigation measures.

On the day the data breach was reported (18 May 2017), the firm reset the passwords for all affected users, log them out of the app and website, and agreed to provide monetary rewards to a bug bounty program such that hacker agreed to take the data off the dark web marketplace and delete all copies of them. To put simply, a bug bounty program is an invitation extended by a company to security researchers and potential hackers for them to point out prospective security flaws or bugs in its system, in return for some rewards. We note that the bug bounty program was not new to the company. It was launched in February 2016, more than a year before the May 2017 data breach; the incentives for potential hackers to report vulnerabilities were perceived to be limited. For instance, the initial report of the May 2017 data breach by HackRead stated "It must

be noted that the breached firm [firm name masked] already has an existing bug bounty program however the security researchers and hackers who report vulnerabilities only receive Hall of Fame recognition or a certificate of acknowledgment."

To investigate whether the firm kept its promise and increased the rewards for reporting security flaws after the data breach, we reviewed the history of the firm's changes in its bug bounty program policy, which is available in the program's website. On June 28, 2017, the firm adopted a monetary reward system that will reward reports according to their severity on a case-by-case basis as determined by the firm's own security team, with a minimum reward for severe bugs such as "Remote Code Execution" or "User Personal Information Access" being $1000 USD. In the next two weeks, the firm added restrictions for claiming monetary rewards (e.g., "We only reward the first reporter of a vulnerability. Public disclosure of the vulnerability prior to resolution will result in disqualification from the program. You must report a qualifying vulnerability through the HackerOne reporting tool to be eligible for a monetary reward.") and extended the list of non-qualifying vulnerabilities.

In sum, after the incident of the data breach, the breached firm persuaded the hacker to delete the breached data and introduced monetary incentives for reporting security flaws in its existing bug bounty program. Such measures signal that the firm is taking actions to reduce future breach risks. The measures on 18 May 2017 (resetting passwords, persuading hackers, and announcing monetary rewards for a bounty program) are unlikely to drive our results as they occurred on the day of the breach announcement, while we observe gradual recovery of digital payment in the several weeks afterwards. Nevertheless, the subsequent implementation of the monetary incentives for the existing bounty program might change customers perception. Even though it had been announced on 18 May 2017 and is subject to restrictions, we do not tend to make a strong claim that it does not affect consumers' perception at all. Therefore, we acknowledge that while the common occurrence of repeated data breaches can lead consumers to be concerned about future data breaches, the measures by the breached firm may partially mitigate consumers' concerns about future data breaches.

A.4. The Data Breach on the Online Grocery Store

A.4.1 Data Breach and Sample Construction

We also perform analyses on the data breach associated with two other samples. All the samples in this paper are from India which is the fourth-worst in terms of the time taken to identify and contain a data breach. Thus, data breaches are a critical concern in India.[5]

The first one is a leading online grocery store in India. Through its online platform, consumers order groceries that will be delivered to them. For the online grocery store, there was a data breach on Facebook on September 25, 2018, and hackers could "take full control of the victim's account, including logging into third-party applications that use Facebook Login."[6] Consumers can opt to use Facebook to enable login to the online grocery store, and save their card details to make frequent payments. Therefore, the Facebook data breach was a concern to the users of the online grocery store.

The online grocery store dataset contains micro-level information about grocery ordering and delivery from January 1, 2016, to April 30, 2019, in Bangalore, Chennai, Gurgaon, Hyderabad,

[5] See https://www.livemint.com/news/india/new-niti-framework-may-help-improve-india-s-data-protection-standards-11600154978292.html (last accessed on September 17, 2020).

[6] See https://www.theguardian.com/technology/2018/sep/28/facebook-50-million-user-accounts-security-berach (last accessed on June 7, 2020).

Mumbai, and Pune. Each customer account in this ordering platform has a unique masked user ID, and we observe the city, date, and value of every order made by each account. For every order, we observe the payment method and the name and quantity of each food item. We consider all non-cash payments as digital payments.

Similar to the food delivery platform sample, we first sort the original order-item level data by payment mode, week, and city. Then we collapse the sample by calculating the number of orders for each payment mode in each week in each city. For this panel dataset, we construct its variants (in terms of frequency and dependent variables) for robustness tests. We examine the differences between digital and cash payments responses to the data breach in a DID framework introduced in Section 3.1. We also investigate quits and payment switch using the method introduced in Section 3.3. The consumer entries cluster at the beginning of the sample period, January of 2016, and we are not able to explore changes in consumer entries.

A.4.2 Results for the Online Grocery Store Sample

In Figure A.8, we construct seasonality-adjusted daily times series of order count, consumer count, and sales for both digital and cash payments. The method is similar to the one in Section A.1.1:

$$\text{Ln}(Y_{pt}) = \beta_0 + \lambda_{pd} + \mu_t + \gamma_{pm} + \varepsilon_{pt}, \qquad \text{(A.6)}$$

where p, t, and d represent payment mode, date, and day of week, respectively. Y_{pt} is the number of orders using payment mode p on date t. λ_{pd} is the payment mode - day of week fixed effects. $\boldsymbol{\lambda_{pd}}$ controls for the within-week cyclicality which is different between digital and cash payments. μ_t controls for the date fixed effects. We also find strong within-month cyclicality. Thus, we include an additional term γ_{pm}, the payment mode - day of month fixed effects. Similar to the food delivery breach, we drop observations before April 1, 2017, to eliminate the effects of the demonetization in India. We then use the residuals $\boldsymbol{\varepsilon_{pt}}$ as the time series of daily order count/user count/sales, one for digital payments ($\boldsymbol{\varepsilon_{pt}}$ where p is digital payment), the other for cash payments ($\boldsymbol{\varepsilon_{pt}}$ where p is cash payment).[7] We only find mild declines in digital payments on the day of the data breach in the time series of order count and consumer count and little change in sales.

We then investigate the change in payment rigorously using the DID regression similar to model (2):

$$\text{Ln}(1+Y_{ptz}) = \beta_0 + \beta_1 \, Treat_p \times T_t + \beta_2 \, Treat_p + \lambda_{tz} + \varepsilon_{ptz}, \qquad \text{(A.7)}$$

where p, t, and z represent payment modes (digital or cash), weeks, and cities, respectively. Note we do not observe subzones in the online grocery store sample. Y_{ptz} is the number of orders paid by mode p (cash or digital payment) in week t in city z. In the robustness tests, we also consider daily frequency. $Treat_p$ equals 1 if p is the digital payment and equals 0 if p is the cash payment. T_t equals 1 if week t is equal to or later than the 39th week of 2018 in which the breach was announced and equals 0 otherwise. λ_{tz} is the week-city fixed effects that control for time-varying shocks within each city. λ_{tz} controls for the time trend and allows heterogeneities across different cities. The three terms constitute a DID specification. Robust standard errors are clustered at the city level. We restrict the sample to the period from week 35 to week 42, i.e. four weeks before and four weeks since week 39. Table A.6 displays the results, and estimates are

[7] To make the pattern more clear, we use the seven-day moving average (day t - 6 to t) of the residual to plot for all daily samples in this paper except those for Figure 1, Figure A.2, Figure A.5, and Figure A.6 in which the trend is clear.

positive and statistically and economically insignificant (0.059 and 0.049), namely, there is little change in payment choice.

We also analyze user quits in a similar way to those in Section 3.3 and Table 3 of the main text. The week of a quit is defined as the first week of four-week inactivity. Figure A.9 shows that user quits did not increase after the data breach. The results in Columns 1 and 2 in Panel A of Table A.7 indicate that users with higher digital payment ratios before the breach did not quit more after the breach as the marginal effect of *Digital_ratio* is only -0.004 and statistically insignificant.[8] However, Columns 3, 4, 5, and 6 show that digital payment users tend to switch from digital payment to cash after the breach.

A.5. The Data Breach on the Bank

A.5.1 Data Breach and Sample Construction

Second, we investigate a data breach in the Indian banking system that made banks ask their customers to change the security codes of as many as 3.25 million debit cards, which was reported on October 19, 2016.[9]

We obtained a data from a leading Indian Bank. The bank dataset contains micro-level information about account opening, credit limit change, transactions using bank cards from January 1, 2014, to December 31, 2017, in Ahmedabad, Bangalore, Bhubaneshwar, Chennai, Delhi, Gurgaon, Kolkata, Mumbai, and Surat. We observe the masked customer ID, the rupee amount of transaction, card type (debit card or credit card), city, account opening date, credit limit change date, and the amount of credit limit.

We first sort the original transaction-level data by date and card type. We then collapse the sample by calculating the number of and the total rupee amount of transactions for each card type on each day. We compare the rupee amount and the number of transactions before and after the breach for credit card and debit card respectively.

A.5.2 Results for the Bank Sample

The subsection explores the effects of the data breach of the Indian banking system on bank users. We exploit the sample constructed in Section A.5.1 on the number of and the total rupee amount of transactions for each card type on each day. We consider the sample between 19 July 2016 (two months before the report date) and 8 November (the day before the demonetization).[10]

[8] There would be only 118 and 80 observations in Column 1 and 2 respectively if since we control for city fixed effects, as the dependent variable is a dummy. Therefore, we do not control for city fixed effects in the two columns. Additionally, if we control for account age in Column 2, there is no variation in the dependent variable for all the observations except when the value of account age is 142 (days), so all of these observations (with account age's value inequal to 142) are dropped, and we still cannot get estimate for account age since it has only one value in the final sample. Thus we do not control for account age in Column 2.

[9] We further discuss timeline of this breach in Footnote 10 of this appendix.

[10] The breach was believed to have occurred since May 2016. However, it was only in early September that banks and payments services providers became aware of the extent of the breach. News reports by mainstream media about the breach occurred mostly after October 19, 2016. Therefore, it is possible that banks took actions earlier and some bank customers were aware of this breach before those reports. Thus, it is difficult to identify a precise event day similar to the one for the food delivery platform. We consider October 19 as the event day since media reports can be considered as an additional shock to consumers, even if some of them were already aware of the breach.

We first seasonally adjust the time series for credit card and debit card respectively using the following regression:

$$\text{Ln}(Y_t) = \beta_0 + \lambda_d + \varepsilon_t, \qquad \text{(A.8)}$$

where t and d represent date and day of week respectively. We first restrict the sample to a certain card type and run the regression. Y_t is the number of transactions or the rupee volume of transactions using that type of card on date t. λ_d is day of week fixed effects. λ_d controls for the within-week cyclicality. We then use the residuals ε_{pt} as the time series of daily transaction count/rupee volume. We plot the time series in Figure A.10. Before and after October 19, 2016, there was little change in the transaction count and transaction rupee volume for both credit cards and debit cards.

Conceptually, the data breach can be viewed as a negative shock to consumers' uncertainty about the future availability of liquidity/cash. D'Acunto et al. (2020) show that users of a digital bank who have the largest amount of liquidity through deposits increase their spending by more than for others after a shock to the uncertainty of future liquidity. Motivated by this, we also consider heterogeneity due to wealth effects. In other words, different consumers with different levels of wealth may respond differently, which is masked by the aggregate effects. We use asset balance on bank account and credit limit to proxy wealth, respectively. In particular, the account balance contains the balance of all assets of a consumer in that bank. The values of both the account balance and the credit limit are from the last records before the bank breach. In particular, the account balance does not necessarily include all wealth of the consumer, but it is reasonable that the account balance is correlated with the total wealth of the consumer.

We first categorize consumers into to a wealthier group whose account balance is higher than the median wealth and a less wealthy group whose account balance is lower of equivalent to the median wealth. We then construct two daily-group level panel samples, one for credit card, and the other for debit card. In the construction of these two samples, we calculate the number of and the rupee amount transactions (of a card type) for the wealthier group and the less wealthy group respectively. We seasonally adjust two samples separately using the following regression:

$$\text{Ln}(Y_t) = \beta_0 + \lambda_d + w + \varepsilon_t, \qquad \text{(A.9)}$$

where t and d represent date and day of week respectively. We first restrict the sample to a certain card type and run the regression. Y_t is the number of transactions or the rupee volume of transactions using that type of card on date t. λ_d is day of week fixed effects. λ_d controls for the within-week cyclicality. w is the group fixed effects. We then use the residuals $\boldsymbol{\varepsilon_t}$ for one of the two groups as the time series of daily transaction count/rupee volume for that group. We plot the two time series in Figure A.11. The solid line represents the wealthier group, and the dashed line represents the less wealth group. Before and after October 19, 2016, there was little difference in the trend between the two groups for both credit cards and debit cards. We also use credit limit as the proxy for wealth and repeat the above process. The results are in Figure A.12 and indicate little difference in the trends of the two groups.

A.6. Further Tests and Discussions

A.6.1 Power Analyses

The largely null results raise a concern of whether the effect is close to 0 or the data does not have enough statistical power. If the statistical power, i.e., the probability of avoiding a Type II error (false negative error), is found to be low, one would regard the null results as ambiguous,

since failure to reject the null hypothesis cannot have much substantive meaning when the probability of rejecting the null hypothesis was low even though the phenomenon exists.

We first conduct power calculations for our setting. Different from the typical ex ante power calculations to evaluate the trade-off between sample size and statistical precision in designing experiments, our power calculations are *ex post* in nature: after obtaining insignificant regression results, we go back to gauge the statistical power in our settings. Details of the power calculation (based on notations from Burlig, Preonas, and Woerman (2020)) are as follows.

Power calculations provide an *ex ante* estimate of the smallest effect size that an experiment, with a given sample size and experimental design will be able to statistically detect. Typically, power calculations take the following form:

$$\text{MDE} = \left(\text{t}^{\text{d}}_{1-\kappa} + \text{t}^{\text{d}}_{\alpha/2}\right)\sqrt{\text{Var}(\hat{\tau}|\mathbf{X})}$$

where $\text{Var}(\hat{\tau}|\mathbf{X})$ is the exact finite sample variance of the treatment effect estimator, conditional on independent variables $\mathbf{X}$; $t^d_{\alpha/2}$ is the critical value of a t distribution with d degrees of freedom associated with the probability of a Type I error, α, in a two-sided test against a null hypothesis of $\tau = 0$; and $t^d_{1-\kappa}$ is the critical value associated with the probability of correctly rejecting a false null (a Type II error), κ. The degrees of freedom, d, will depend on the dimensions of $\mathbf{X}$ and the treatment effect estimator in question. These parameters determine the minimum detectable effect (MDE), the smallest value $|\tau| > 0$ for which the experiment will correctly reject the null $\tau = 0$ with probability κ at the significance level α.

In difference-in-differences research designs, non-constant within-unit serial correlation in panel data not only affects statistical inference (Bertrand, Duflo, and Mullainathan, 2004)**,** it can also impact statistical power in a similar way (Frison and Pocock, 1992). We use the "serial-correlation-robust" (SCR) power calculations for the difference-in-differences (DD) estimator developed by Burlig, Preonas, and Woerman (2020)'s accompanying Stata package, pcpanel, to perform the power calculations.

We parametrize analytical DD power calculations by directly estimating the idiosyncratic residual variance and covariances of the outcome variable in the actual transaction data. The resulting power calculations can capture the complex dependencies inherent to real data. The numbers of cross-sectional units, pre-treatment periods, and post-treatment periods, which are required in the SCR DD power calculations, are set to be equal to the corresponding numbers in our regression settings.

Following the norm of calculating the statistical power to choose the sample size in experimental studies, we aim for 80% power (κ). Under the default $\kappa = 80\%$ and $\alpha = 5\%$, we calculate the MDEs according to the flexible SCR DD power calculations. As a comparison, we also calculate the MDEs according to the more restrictive McKenzie (2012) power calculations.

For panel regressions where we include fixed effects, we first de-mean the raw level of the dependent variable with respect to all layers of fixed effects in regression equations. We then calculate the standard deviation of the residualized dependent variable.

We conduct the above power analyses for all regressions with insignificant estimates for the key independent variables (such as Digital × Post). We tabulate the results in Table A.8. For all but three of our insignificant estimates, the results do not support the argument that the null results are due to low statistical power. For instance, for the weekly analysis in Table 1 Panel A Column 2, the SCR power calculations suggest an MDE that is economically small, only 0.08 of standard deviation of the dependent variable. In other words, our sample size is large enough to detect an

effect as small as 0.08 of standard deviation of the dependent variable when we aim for 80% power. For the sake of comparison, Cohen (1988) considers an effect size below 0.2 as a small effect. From these calculations, we conclude that these null results are not driven by insufficient statistical power.

The exception is Column 1 of Table A.6 whose MDE is bigger than 0.2 of standard deviation. But the insufficient power is not a serious concern because when we use the daily samples with sufficient statistical power in Column 2 of Table A.6, we still do not find statistically declines in digital payments.

Lastly, we would also like to point out that a few factors that boost the statistical power in our settings. First, the data were collected directly from the apps' transaction records, and had very little (if any) measurement error. Second, we use weekly or daily data in our analysis, which provided high-frequency observations to detect the effect of data breaches. McKenzie (2012) shows that panel data experiments, where multiple observations per unit are taken, help to average out noise and thus increase statistical power.

A.6.2 External Validity and Implications for Data Protection Policies

Our findings are based on datasets from India. For the following reason, we believe that consumers in other countries are likely to behave similarly to our findings.

First, conceptually, the effect of a data breach on consumer behaviors crucially depends on the nature of the breach. A data breach that compromises sensitive personal information such as Social Security numbers or bank account numbers is likely to have a larger impact on consumer behaviors than a more minor breach. To better interpret the magnitude of our estimated effects, we compare the type of data involved in the specific data breaches we study with a larger and more representative sample of data breaches compiled by Internet security service Have I Been Pwned (https://haveibeenpwned.com/). The service tracks data breaches by collecting and analyzing the leaked data dumps and reports what specific types of data were included in each of the data breaches it monitors. Compared with the Privacy Rights Clearinghouse that we use in our analysis of repeated data breaches, Have I Been Pwned tracks a smaller number of data breaches and covers a more recent time period. One crucial advantage is that this data set includes information on what specific types of data were included in each of the data breaches. We analyze the nature of leaked data in the 577 data breaches that took place during the 11 years from 2011 to 2021. We define "sensitive personal information" as financial information (bank account, credit card, financial investments, financial transactions) or personal identification information (Social Security numbers, government issued IDs). Using this definition, we find that only 42 or 7.3% of the breaches involved leakage of some sensitive personal information; the other 535 or 92.7% of the breaches did not involve leakage of any sensitive personal information. In other words, similar to the data breaches that we analyze, the majority of data breaches do not involve leakage of sensitive personal information.

Within the scope of our analysis, we focus on the comparison of short-term and long-term effects. We find a transient impact of the data breaches where usage of digital payments drops initially but recovers in the longer term. While we cannot directly infer whether one would find similar effects from other data breaches, the similarities that the data breaches in our samples have with other data breaches support the generalizability of our findings. We acknowledge that for more serious data breaches like those that involve sensitive personal information, even the long-term effects could be more pronounced.

Second, existing studies in household finance document that Indian consumers behave similarly to those in other countries. For example, D'Acunto, Prabhala, and Rossi (2019) find that India consumers exhibit similar behavior bias in financial decision making. On the issue of personal data sharing and data protection more specifically, findings from several experimental and survey-based studies imply that consumers' response to a breach of their information is likely to be modest in the U.S. context. Athey, Catalini, and Tucker (2017) conduct experiments in the U.S. and find that even people who claim to value privacy are willing to relinquish their private data for small benefits. Based on a 2015 nationally representative survey on consumer attitudes towards data breaches in the U.S. conducted by the RAND corporation, only 11% of data breach victims stop dealing with the company following a breach (Ablon et al., 2016). We note that the U.S. setting, generally more familiar to and well understood by academic researchers, is characterized by better data protection and stronger emphasis on privacy.

Third, contrary to a setting where substantial centralization of economic agents' personal data exists, such as China, the India setting resembles those of other countries in this respect. This similarity bolsters the generalizability of our findings. Moreover, India's media sector can mention potential data breaches freely, and hence consumers are likely to be informed about these issues in the same way as consumers in most other countries are informed.[11]

In our findings, the declines in digital payments vanished in the long run. We acknowledge that for data breaches that are more serious than those we study, even the long-term effects could be significant. We focus on the difference between the short term and the long-term effects, namely, the long-term reaction is smaller than the short-term reaction.

In recent years, growing concerns for consumer privacy have prompted the enactment of data protection laws such as EU's General Data Protection Regulation (GDPR). While we do not directly study regulatory policies, our analyses have implications for the policy debate on consumer data protection. An important objective of data protection laws is to restrict collection of personal data, which can be leaked in cybersecurity breaches or misused for unwarranted reasons (e.g., discrimination, exploitation of consumer vulnerability). These undesirable costs need to be weighed against the improved access to products and services and the substantial convenience enabled by big data technologies. In this study, we focus on the setting of data breaches, a risk factor from which data protection laws aim at protecting consumer, to analyze this trade-off. When the information leakage risk materializes in a data breach, consumers may or may not change usage patterns, depending on how much utility they gain from the convenience relative to the disutility of the information leakage. By inferring consumer preferences from observed behaviors in unexpected data breaches, we empirically demonstrate that for consumers, the benefit of convenience outweighs the cost of information breaches. Our findings are related to Liu, Sockin, and Xiong (2021)'s theoretical framework: For temptation goods such as gambling or video games, consumers hide private information to concel behavioral vulnerabilities; for normal consumption goods such as food delivery, grocery, and banking service that we study in this paper, consumers share data for convenience.

In a similar vein to our study, a number of survey and experimental studies show a paradox that consumers' self-reported preferences for sharing/protecting personal data are inconsistent with their actual behaviors, a phenomenon referred to as the "privacy paradox" (see Acquisti, Brandimarte, and Loewenstein (2020) for a summary). Although we are unable to directly measure consumer perceptions from actual behaviors, our approach of analyzing consumer activity has one

[11] We thank an anonymous referee for this suggestion.

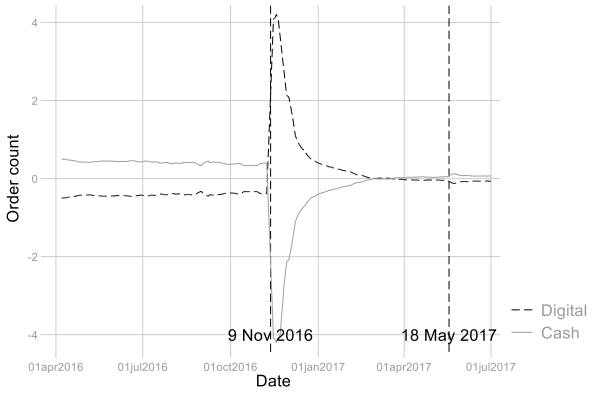
key advantage – it does not rely on participant self-report and recall and therefore is not subject to being incomplete due to limits of memory, under-reported due to forgetting past experiences, or over-reported due to being influenced by media coverage. Therefore, our study complements these survey and experimental studies to provide additional empirical estimates for consumers' information sharing preferences and how their preferences affect their data-sharing choices.

A.6.3 Vegetarianism in India

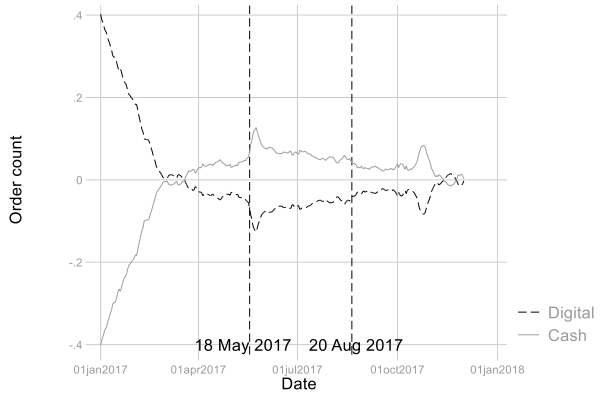
According to the statistics from UN Food and Agriculture Organization, India's meat consumption per person was lowest among 177 countries globally in 2007. Vegetarianism in India is prevalent and can be influenced by religious, cultural, and socioeconomic factors. We would like to clarify that for our analysis, we analyze the vegetarian index to examine whether the composition of new customers changes before and after the data breach on the food delivery app and we remain agnostic to the underlying cause(s) of vegetarian diet. Reassuringly, our transaction-based measure of the proportion of vegetarians (36% both before and after the data breach) is broadly consistent with survey-based measures from national surveys including India's official Sample Registration System Baseline Survey 2014, the 2015-16 India National Family Health Survey (NFHS-4), and a 2021 survey conducted by Pew Research Center.

Figure A.1 Demonetization

Panel A Long term

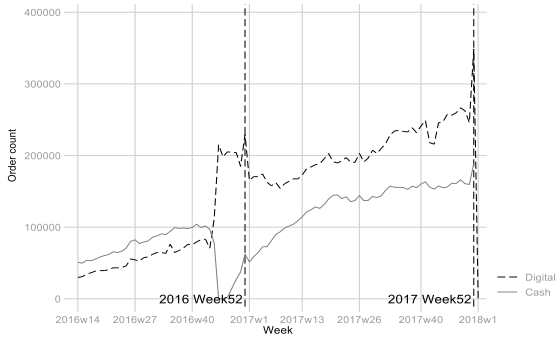


Panel B Short term



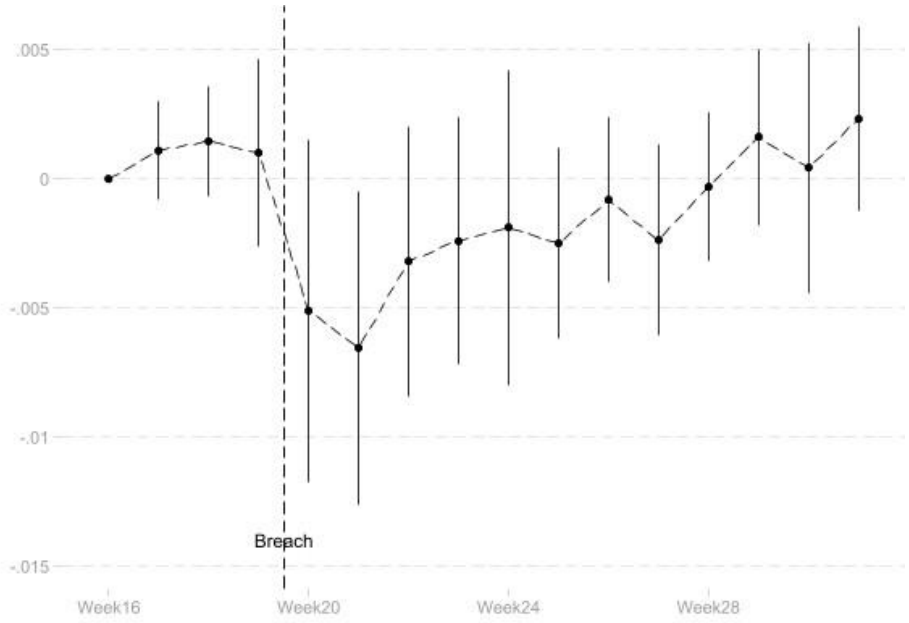
The figure shows the daily time series of the total number of orders paid by digital payments and cash payments respectively. The times series are seasonally adjusted by the method described in Section A.1.1. We add one to the dependent variable before taking the logarithm in the seasonality adjustment. In the step of seasonality adjustment, 2 out of 1282 observations have zero values for the dependent variable.

Figure A.2 Year-end spikes



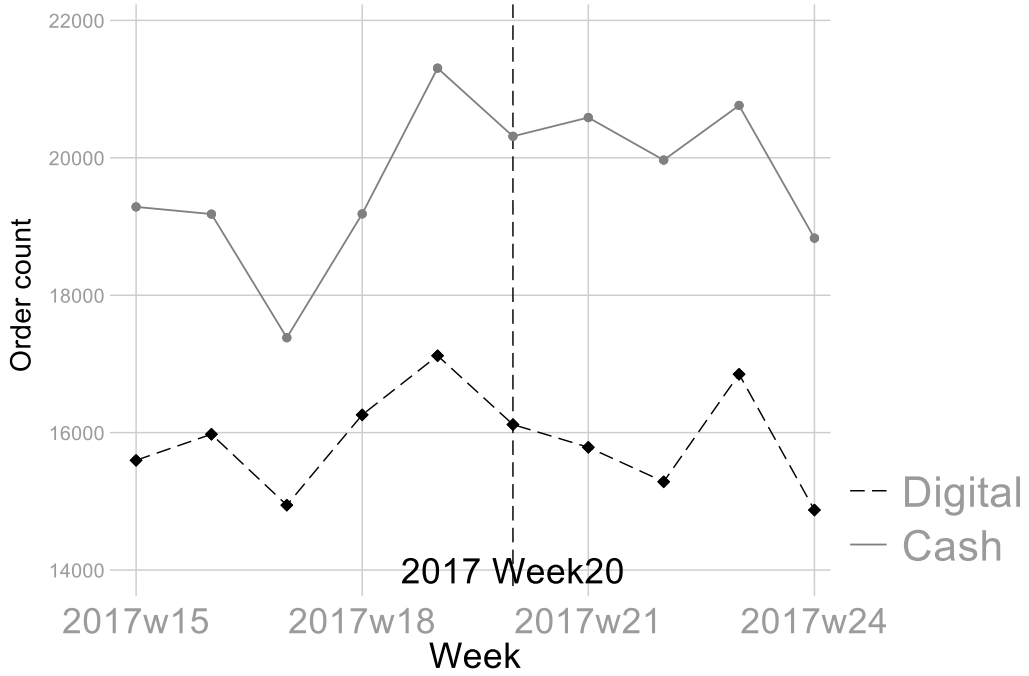
The figure shows the weekly time series of the total number of orders paid by digital payments and cash payments respectively.

Figure A.3 Event Study



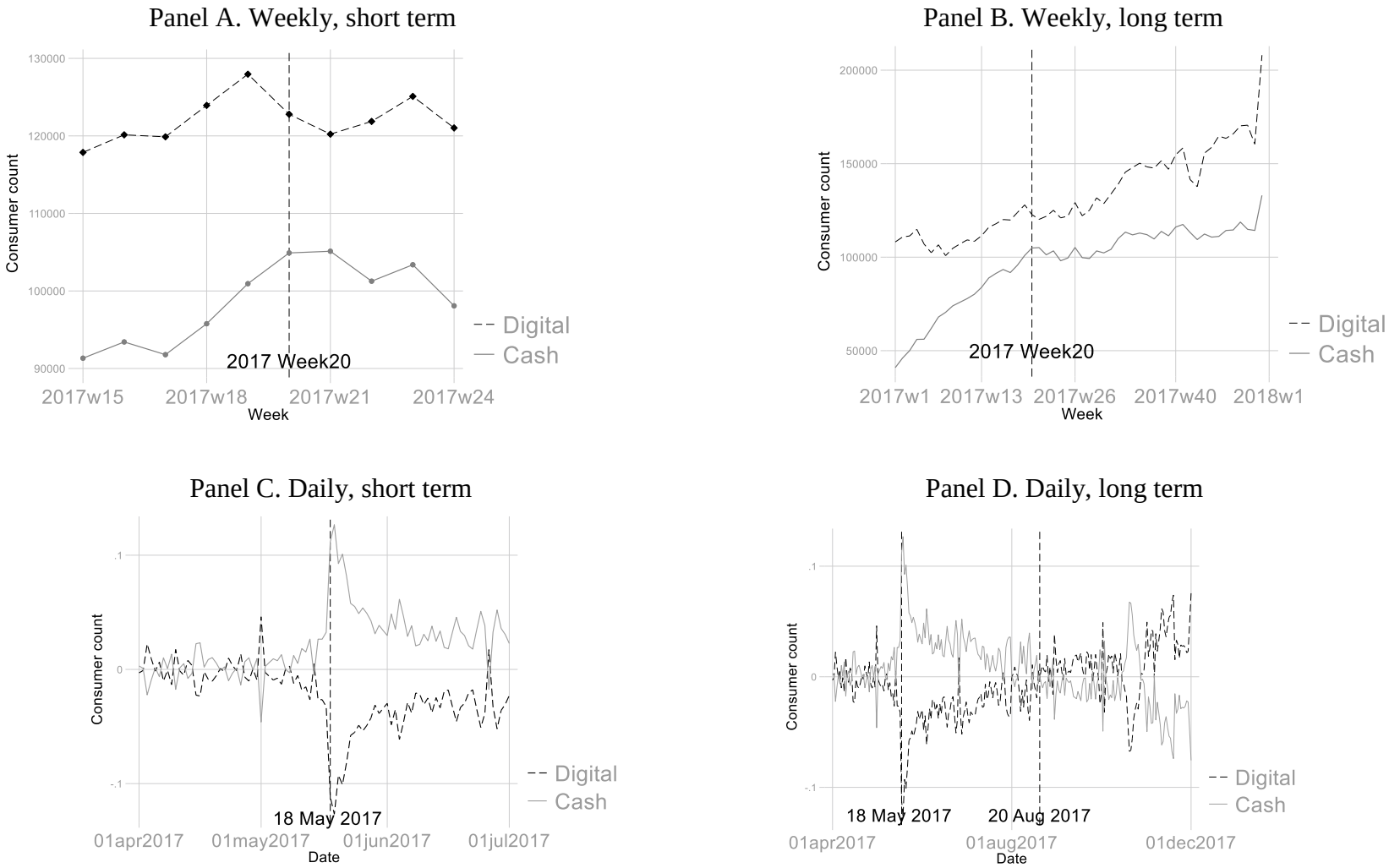
The figure plots the point estimates and the 95% confidence intervals for the coefficients in the event study using the food delivery platform sample. Note that the breach was in Week 20.

Figure A.4 Entrants: Cash and Digital Payments



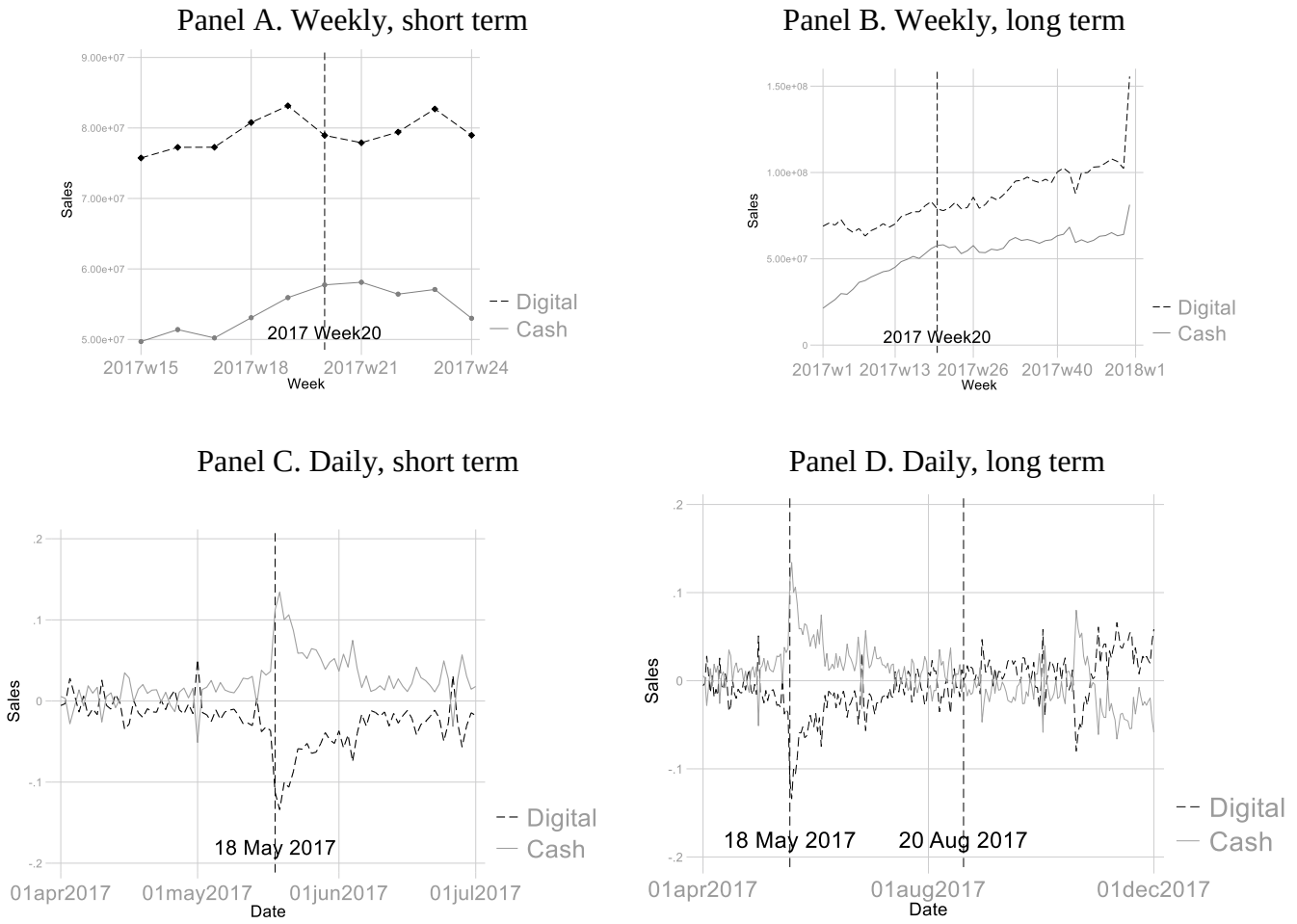
The figure shows the weekly time series of order counts for cash and digital payments by entrants (i.e. during the week they appear in the platform for the first time).

Figure A.5 Changes in Payment Modes - Consumer Count



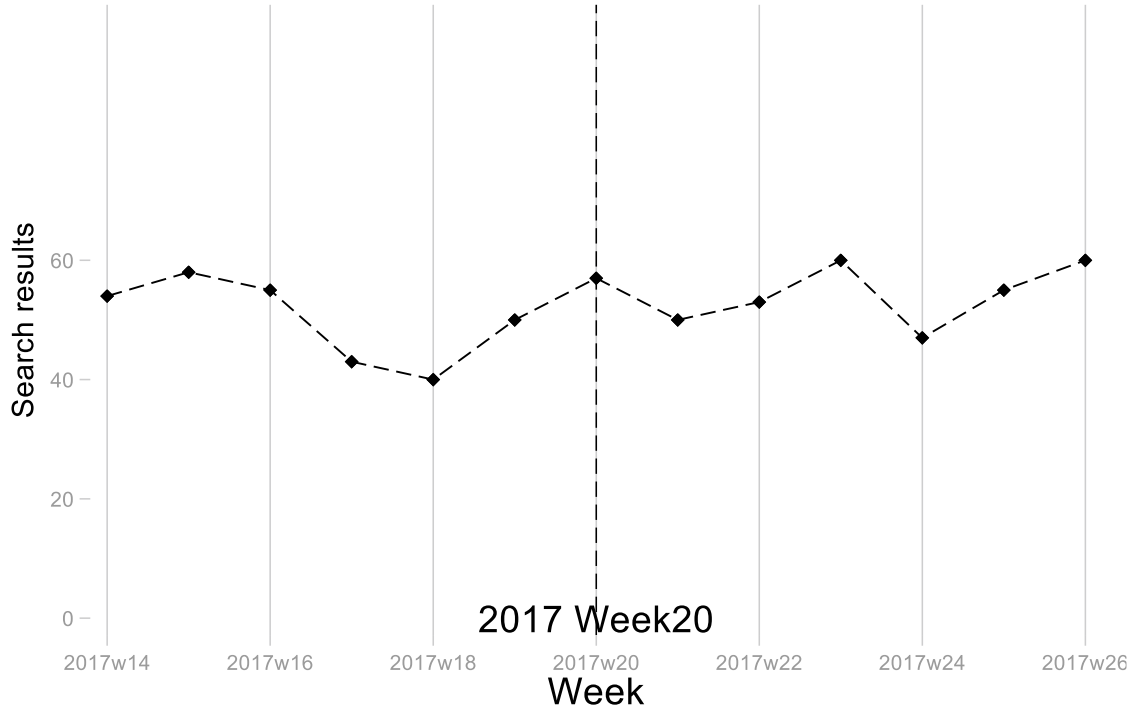
The figure shows the time series of the number of consumers who use digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample, and Panels C and D are for the daily sample. The daily sample is seasonally adjusted as mentioned in Section A.1.1.

Figure A.6 Changes in Payment Modes – Sales



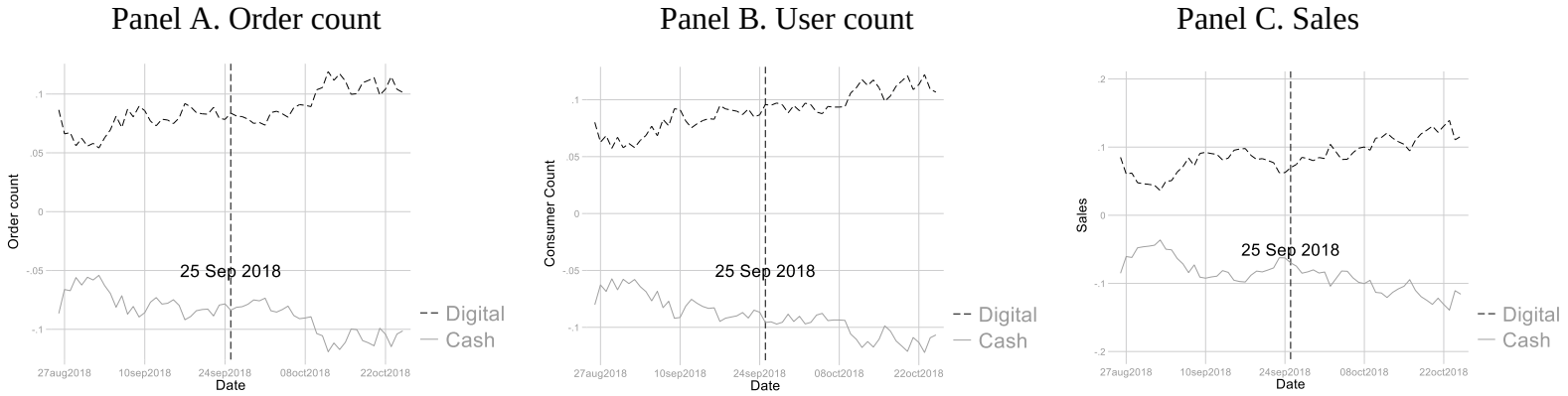
The figure shows the time series of the rupee volume of sales for digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample, and Panels C and D are for the daily sample. The daily sample is seasonally adjusted as mentioned in Section A.1.1. The calculation of sales is introduced in Section A.2.5.

Figure A.7 Time Series of Promotion



The figure shows the weekly time series of the number of search results associated with the platforms' promotion. Details of the search process is described in Section A.2.6.

Figure A.8 Grocery Store: Changes in Payment Modes

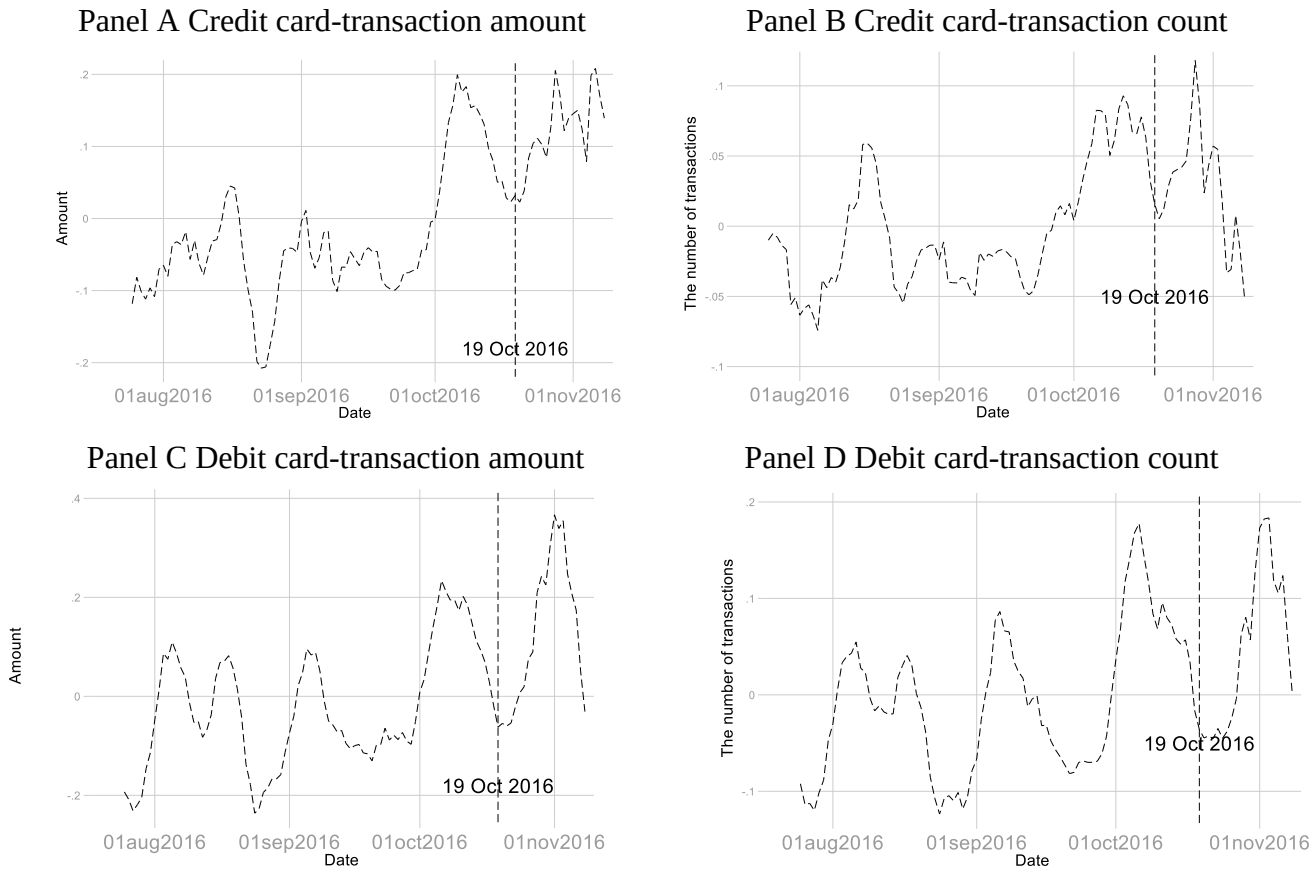


The figure shows the daily time series of the user count, order count, and the rupee volume of sales for digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. The time series is for the sample of the online grocery store and is seasonally adjusted as mentioned in Section A.4.2.

Figure A.9 Grocery Store: User Quits

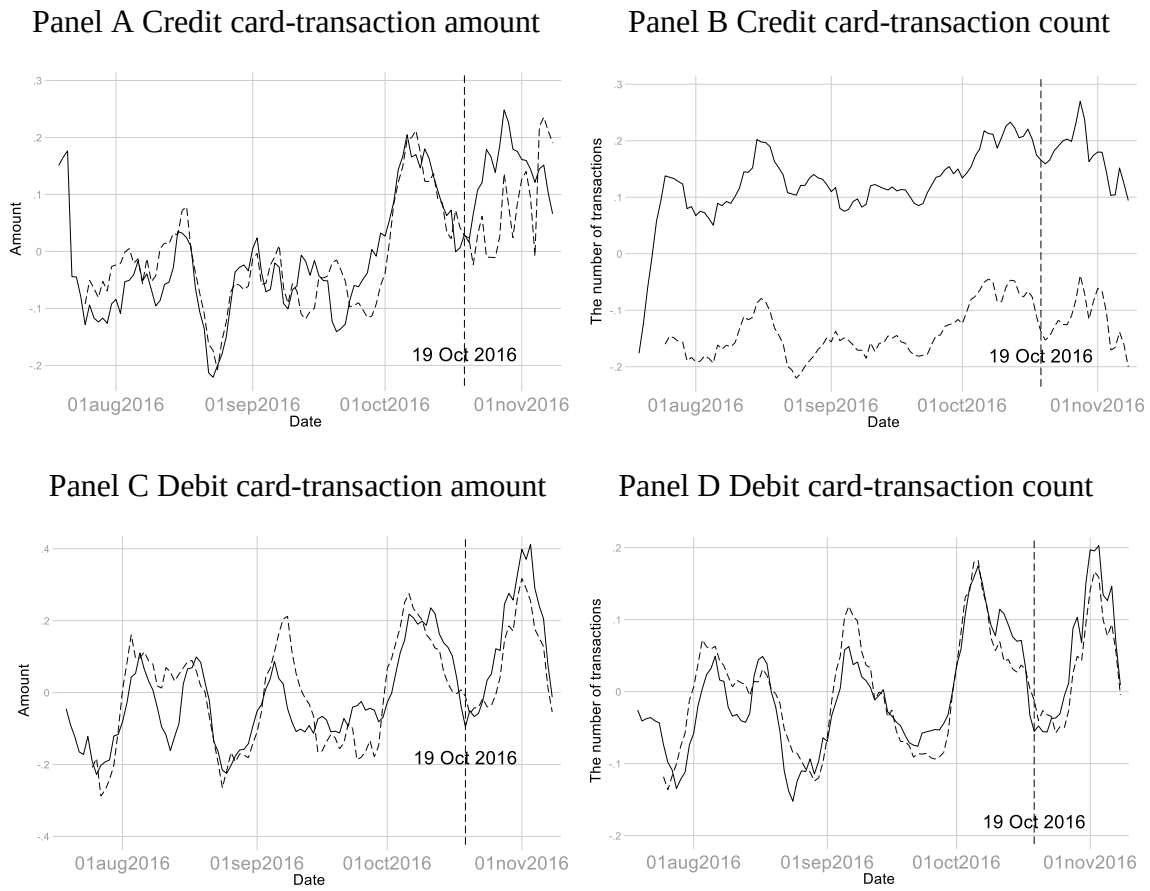
The figure shows the weekly time series of consumer quits. The sample is for the online grocery store. The week of a quit is defined as the first week of a four-week inactivity.

Figure A.10 Transaction Change of Bank Users: the Data Breach in Indian Banking System



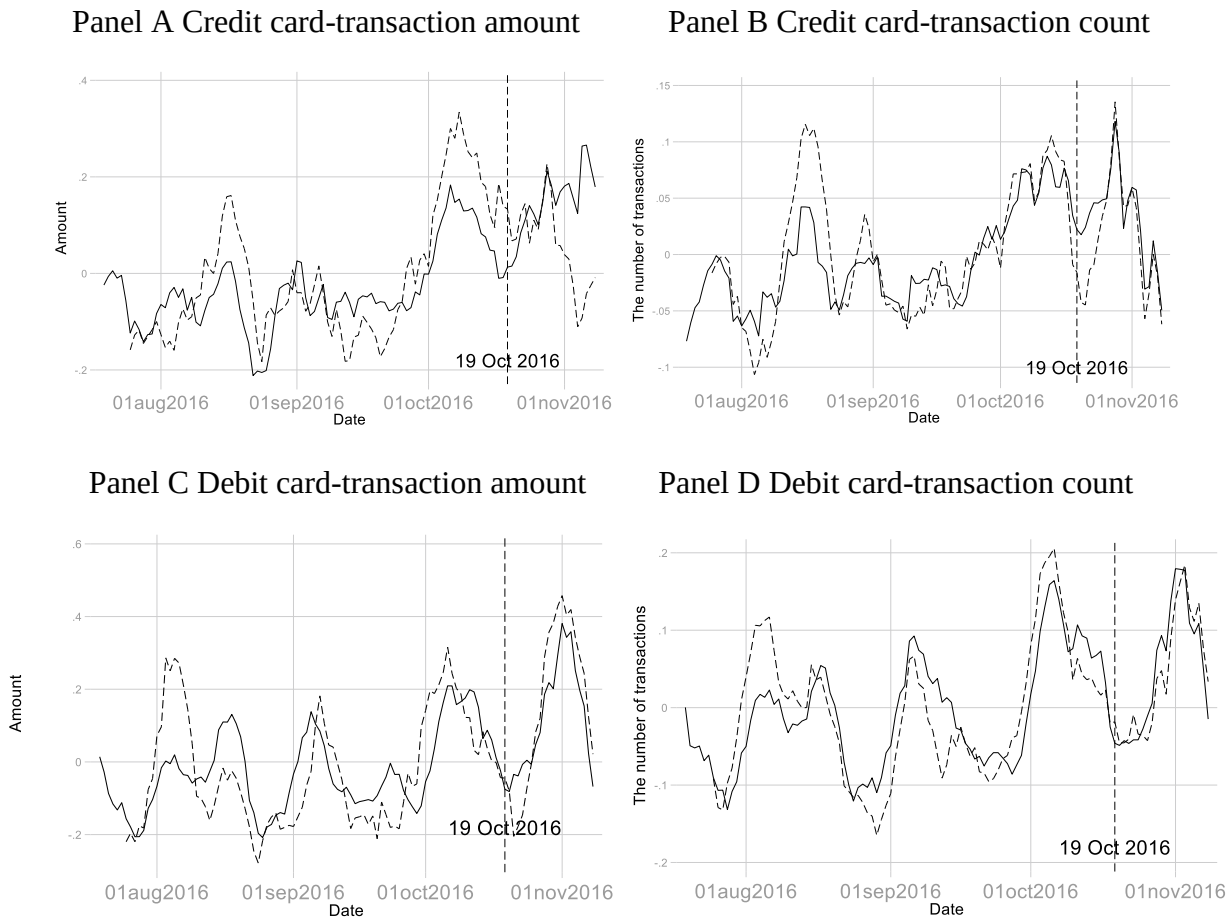
The figure shows the effects of the bank data breach on the transactions of the bank users. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Figure A.11 Transaction Change of Bank Users: the Data Breach in Indian Banking System – by Account Balance



The figure shows the effects of the bank data breach on the transactions of the high–wealth and low-wealth bank users defined in Section A.5.2. The solid line represents high-wealth users, and the dash line represents the low-wealth users. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Figure A.12 Transaction Change of Bank Users: the Data Breach in Indian Banking System – By Credit Lmit



The figure shows the effects of the bank data breach on the transactions of the bank users with high and low credit limit defined in Section A.5.2. The solid line represents users with high credit limit, and the dash line represents the users with low credit limit. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Table A.1 Daily Analyses: Change in Payment Modes for the Food Delivery Platform Sample

Dep. var.: Ln(1+order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.060***	0.016
	(-6.17)	(1.40)
Payment Mode FE(Digital dummy)	Yes	Yes
Date × Subzone FE	Yes	Yes
Observations	79112	77836
Observations with DV=0	13052	11208
R^2	0.964	0.963
Mean DV	2.433	2.495
P value for wild-t	0.054	0.281

Subzone-payment mode-daily level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively, and the dependent variable is the natural logarithm of the number of orders. The mean dependent variable, ln(1+*y*), is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A.2 User-level Analyses: Change in Payment Modes for the Food Delivery Platform Sample

Panel A DiD

	(1)	(2)
Dep. Var.	Ln(1+order count)	Ln(1+order count)-long term
Post × Digital	-0.012***	-0.003
	(-4.81)	(-1.68)
Payment Mode FE	Yes	Yes
User-Week FE	Yes	Yes
Observations	7287168	7287168
Observations with DV=0	5,862,986	5,994,172
R^2	0.510	0.518
Mean DV	0.172	0.156

Panel B Time-series

	(1)	(2)
Dep. var.	Digital payment ratio	Digital payment ratio-long term
Post	-0.013***	0.012***
	(-13.13)	(9.87)
User FE	Yes	Yes
Observations	384,231	347577
Observations with DV=0	NA	NA
R^2	0.756	0.743
Mean DV	0.636	0.653

The table shows regressions estimating the response of order placing and payment modes to the data breach. Panel A is at the user-payment mode-weekly level. Panel B is at the user-weekly level. The sample is restricted to users who ordered at least once during the month right before the breach. In both panels, Column (1) use the short-term period and Columns (2) uses the long-term period. Both sample periods are defined in Section A.2.2. The mean dependent variable, ln(1+*y*), is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the *t*(*G*-1) distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A.3 Extensive Margin and Intensive Margin

	Extensive			Intensive		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Ln(1+order count)	Entrants – short term	Entrants – 4 week window	Entrants – long term	Existing users – short term	Existing users - 2 month	Existing users – long term
Post × Digital	-0.050*	-0.043**	-0.005	-0.063***	0.015*	0.092**
	(-2.76)	(-4.79)	(-0.35)	(-7.23)	(2.39)	(4.19)
Payment Mode FE (Digital dummy)	Yes	Yes	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9920	4960	9920	9952	9952	9952
Observations with DV=0	1214	618	1089	996	935	850
R^2	0.957	0.958	0.953	0.972	0.969	0.970
Mean DV	2.463	2.471	2.471	3.474	3.447	3.455
P value for wild-t	0.167	0.059	0.764	0.033	0.063	0.062

Subsample analyses on extensive margin and intensive margin, respectively. The users for Columns 1 to 3 (entrants) is restricted to the observations for the first week each user uses the platform. Existing users in Columns 4 to 6 are defined as those ordering at least once per week from week 17 to week 20 (and before May 18th). We compare the month before the breach with the first month (Columns 1 and 4) and the third month ((Columns 3 and 6) after the breach respectively. The sample is restricted to a shorter time window (-2 weeks, 2 weeks) in Column 2. In Column 5, we compare the month before the breach with the second month after the breach.

Table A.4 Inattention: DDD using Google Trends

DV: Ln(1+order count)	
Post × Digital × Google search	0.001
	(1.20)
Post × Digital	-0.168**
	(-4.23)
Digital × Google search	0.003
	(1.56)
Post × Google search	-0.001
	(-2.26)
Digital	-0.028
	(-0.20)
Week FE zone FE, clustered at city level	Yes
Observations	10224
Observations with DV=0	944
R^2	0.959
Mean DV	4.081
P value of wild-t for Post × Digital × Google search	0.469

Subzone-payment mode-weekly level regressions exploring how the response of payment modes to the data breach of the food delivery platform is related to Google search index. We compare the month before the breach with the first month after the breach, and the dependent variable is the natural logarithm of the number of orders. The mean dependent variable, ln(1+y), is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A.5 Characteristics of U.S. data breaches (2005-2018)

Total number of data breaches that took place during the period from 2005 to 2018	8,937
Break-down by causes as classified by the Privacy Rights Clearinghouse	
Hacking by an outside party or being infected by malware	2,499 (27.96%)
Unintentional disclosure (sensitive information posted publicly, mishandled or sent to the wrong party via publishing online, sending in an email, sending in a mailing or sending via fax)	1,831 (20.49%)
Physical loss (paper documents that are lost, discarded or stolen)	1,720 (19.25%)
Loss or theft of portable devices (laptops, PDAs, smartphones, memory sticks, CDs, hard drives, data tapes, etc.)	1,171 (13.10%)
Insider (employee, contractor or customer) leakage	606 (6.78%)
Loss or theft of stationary computer (computer or server not designed for mobility)	249 (2.79%)
Fraud involving debit and credit card with skimming devices at point-of-service terminals	68 (0.76%)
Unknown causes – there is not enough information about breach to know how exactly the information was exposed.	793 (8.88%)
Repeated data breaches:	
Total number of companies/organizations	7,573
Number of companies/organizations that experienced repeated data breaches.	940 (12.4%)
Number of companies/organizations that experienced only one data breach.	6,633 (87.6%)

This table displays the characteristics of U.S. data breaches that took place during the 14-year period from 2005 to 2018. Data are from the Privacy Rights Clearinghouse (PRC) chronology of data breaches (https://privacyrights.org/data-breaches). According to the meta data provided by the PRC, this chronology data was last updated on Jan 13, 2020 and as a result, it contains much fewer data breaches in 2019 than in each of previous 5 years. Therefore, we restrict the sample to data breaches that took place from 2005 to 2018.

Table A.6 Changes in Payment Modes for Grocery Store

Dep. var.: Ln(1+order count)	(1) Weekly	(2) Daily
Post × Digital	0.059	0.049
	(1.46)	(1.11)
Payment Mode FE (Digital dummy)	Yes	Yes
Week × City FE	Yes	
Date × City FE		Yes
Observations	96	732
Observations with DV=0	0	0
R^2	0.989	0.979
Mean DV	6.386	4.444
P value for wild-t	0.183	0.286

City-payment mode-weekly (daily) level regressions estimating the response of payment modes to the data breach. We compare the month before and after the Facebook data breach. The dependent variable is the natural logarithm of the number of orders. The mean dependent variable, ln(1+*y*), is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the *t*(*G*-1) distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A.7 Grocery Store: Quits and Payment Switches

Panel A Summary Statistics

Variable	Obs	Mean	Std. Dev.
Quit	815	0.00	0.05
Digital_ratio	815	0.75	0.25
Quantity	815	15.02	7.24
Account Age	815	140.92	8.40
Frequency	815	2.29	1.31
Unit price	815	996.1	502.6
ΔDigital_Ratio	814	0.11	0.19

Panel B Regression Results

Dep. var.	Quit LOGIT Coeff.		ΔDigital_ratio OLS Coeff.		ΔLn(1+digital payments) OLS Coeff.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digital_ratio	-1.474***	-0.848	-0.206***	-0.200***	-1.092***	-1.113***
	(-2.80)	(-1.30)	(-4.46)	(-4.26)	(-6.69)	(-6.96)
Quantity		-0.061***		0.002		0.001
		(-5.79)		(1.97)		(0.27)
Account Age				0.001**		-0.016***
				(3.66)		(-15.01)
Frequency		-2.370***		-0.002		-0.077***
		(-25.21)		(-0.36)		(-9.24)
Price		0.001***		-0.000*		0.000
		(3.50)		(-2.07)		(0.94)
City FE			Yes	Yes	Yes	Yes
Observations	815	815	814	814	814	814
R^2			0.080	0.0840.082	0.205	0.281
Pseudo R^2	0.012	0.151				
Mean DV	0.002	0.002	0.114	0.114	-3.203	-3.203
P value for wild-t	0.452	0.240	0.014	0.014	0.024	0.023
Marginal effects of	-0.004	-0.002				
Digital_ratio	(-1.08)	(-1.83)				

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample is for the online grocery store. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. Panel A displays the descriptive statistics of the sample, and Panel B displays the estimates. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the change in the natural logarithm of the digital payment. Robust standard errors are doubly clustered at the city level. The *t*-statistics are reported in parentheses. The mean dependent variable is reported at the bottom to assess marginal effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A.8 Summary of Power Analysis

Results referenced	Std. dev of DV	SCR MDE	SCR MDE/ std. dev of DV	McKenzie (2012) MDE	McKenzie (2012) MDE/std. dev of DV
Table 1 Panel A Col 2	0.31	0.0368	0.1187	0.0157	0.0506
Table 1 Panel B Col 2	0.12	0.0075	0.0625	0.0066	0.0550
Table 1 Panel B Col 3	0.71	0.1006	0.1417	0.0446	0.0628
Table 1 Panel B Col 4	0.12	0.0093	0.0775	0.0067	0.0558
Table A.1 Col 2	0.31	0.0062	0.0200	0.0019	0.0061
Table A.2 Panel A Col 2	0.25	0.0005	0.0020	0.0004	0.0016
Table A.4	0.32	0.0234	0.0731	0.0123	0.0384
Table A.3 Col 3	0.35	0.029	0.083	0.0159	0.045
Table A.6 Col 1	0.09	0.0479	0.5322	0.0256	0.2844
Table A.6 Col 2	0.20	0.0171	0.0855	0.0093	0.0465

The table shows power analyses for statistically insignificant results.

B. Appendix B: Inverse Hyperbolic Sine Transformation

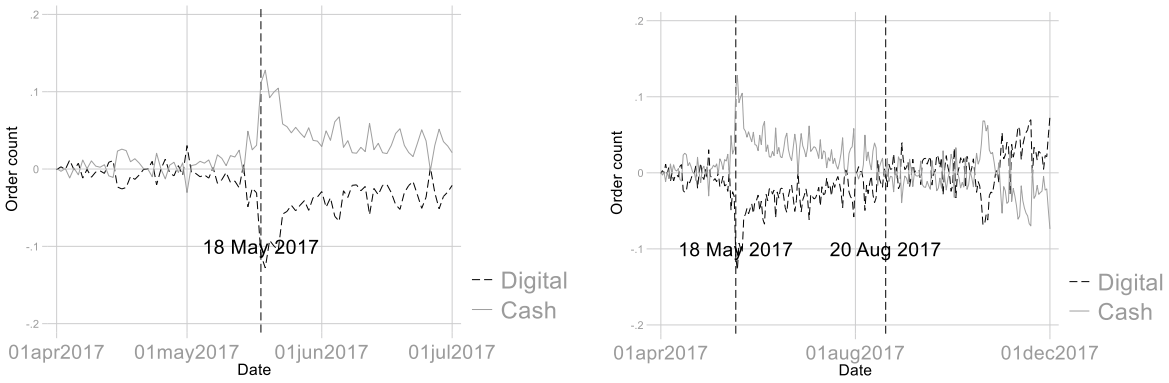
In many previous regressions, we use ln(1+*y*) as the dependent variable for non-negative-valued *y* to retain zero values. Although adding 1 or any other positive constant to a non-negative variable *y* before taking the log retains zero values, it can change the original structure of the data (Duan et al., 1983) and can substantially affect the empirical results (N'guessan et al., 2017). In our sample of week-subzone-payment mode observations, 8.6% of the observations have order count equal to 0. In our sample of date-subzone-payment mode observations, 14.6% of the observations have order count equal to 0. Compared to many other settings used in economics and finance research, the mass of values at 0 in our setting appears smaller: For instance, patent counts are 0 for 69% of Compustat firm-years and toxic release are 0 for 87.6% of establishment-year observations (Cohn, Liu, and Wardlaw, 2021). To investigate the robustness of our results with respect to alternative methods of dealing with zero-value observations, we adopt a commonly used alternative to the logarithm transformation, the inverse hyperbolic sine transformation. Such transformation is a concave log-like transformation and allows retaining zero-valued observations. For a random variable x, taking the inverse hyperbolic sine (arcsinh) transformation yields a new variable $\tilde{x}$ such that $\tilde{x} = \text{arcsinh}(x) = \ln\left(x + \sqrt{x^2 + 1}\right)$. In an "arcsinh-linear" specification where the dependent variable is arcsinh transformed and the explanatory variable is not, the coefficient estimate yields a similar interpretation to that of a standard log-linear specification. See Bellemare and Wichman (2020) for a formal proof.

We repeat all the previous regressions with ln(1+*y*) as the dependent variable and with zero value in the dependent variable using inverse hyperbolic sine transformation. Note the numbers of zero observations for the dependent variable have been reported in the original tables.

For the figure with daily samples and 0 value for ln(1+*y*) as the dependent variable in seasonality adjustment (Figure A.1), we re-estimate the residuals using arcsinh(*x*) as the dependent variable and plot the figures. The numbers of zero value are reported in figures notes in this section. The figures and tables below present the results and show similar patterns to the results obtained with ln(1+*y*) as the outcome variable. Note that we include the original table (figure) numbers in the table (figure) titles for cross reference. Later in Sections D & E, we also use arcsinh(*x*) as the dependent variable in all regressions whose results are tabulated.

Figure B.1 for Figure 1 Panels C and D: Daily Changes in Payment Modes – Order Count

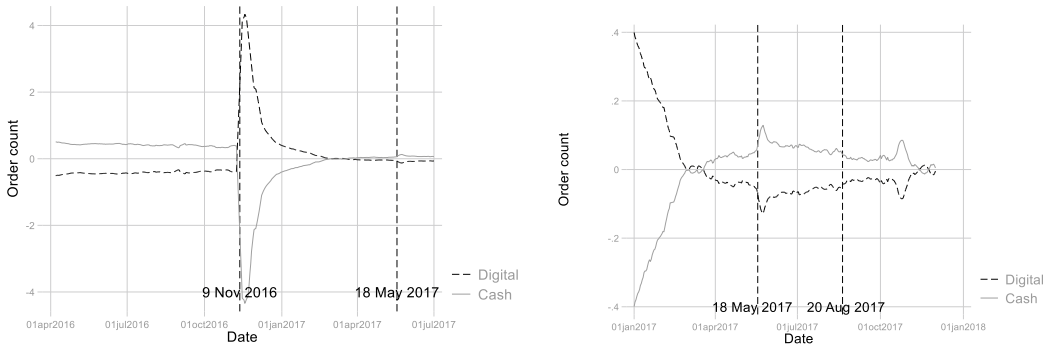
Panel A. Short term **Panel B. Long term**



The figure shows the daily time series of the number of orders paid by digital and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. The times series are seasonally adjusted by the method described in A.1.1 except that the inverse hyperbolic sine transformation is applied to the order count in the adjustment. In the step of seasonality adjustment, 2 out of 1282 observations has 0 value for the order count.

Figure B.2 for Figure A.1 Demonetization

Panel A **Panel B**



The figure shows the daily time series of the total number of orders paid by digital payments and cash payments respectively. The times series are seasonally adjusted by the method described in A.1.1 except that the inverse hyperbolic sine transformation is applied to the order count in the adjustment. In the step of seasonality adjustment, 2 out of 1282 observations has 0 value for the order count before the inverse hyperbolic sine transformation.

Table B.1 for Table 1 Panel A Food Delivery Platform, Weekly

	(1)	(2)
Dep. var.: Arcsinh(order count)	Post: the first month	Post: the third month
Post × Digital	-0.093**	0.024
	(-5.68)	(0.79)
Payment Mode FE (Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	10224	10224
R^2	0.974	0.973
Mean DV	4.670	4.758

Subzone-payment mode-weekly level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table B.2 for Table 1 Panel B Food Delivery Platform, Sales

	(1)	(2)
Dep. var.: Arcsinh(sales)	Post: the first month	Post: the third month
Post × Digital	-0.102**	0.098
	(-3.47)	(1.48)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	10224	10096
Pseudo-R^2	0.949	0.948
Mean DV	10.084	10.281

Subzone-payment mode-weekly level regressions estimating the response of payment modes to the data breach of the food delivery platform. We use the two sample periods respectively and use arcsinh(sales) as the dependent variable. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table B.3 for Table A.1 Food Delivery Platform, Daily

Dep. var.: Arcsinh(order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.068**	0.019
	(-5.19)	(1.28)
Payment Mode FE(Digital dummy)	Yes	Yes
Date × Subzone FE	Yes	Yes
Observations	79112	77836
R^2	0.962	0.961
Mean DV	2.911	2.985

Subzone-payment mode-daily level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table B.4 for Table A.2 Panel A User level Analyses: Change in Payment Modes for the Food Delivery Platform Sample –– Order Count

Dep. var.: Arcsinh(order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.016***	-0.004
	(-4.81)	(-1.69)
Payment Mode FE	Yes	Yes
User-Week FE	Yes	Yes
Observations	7287168	7287168
R^2	0.510	0.519
Mean DV	0.221	0.201

User-payment mode-weekly level regressions estimating the response of order placing and payment modes to the data breach. The sample is restricted to users who ordered at least once during the month right before the breach. Column (1) uses the short-term period and Column (2) uses the long-term period. Both sample periods are defined in Section A.2.2. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table B.5 for Table A.3 Extensive Margin and Intensive Margin

	Extensive			Intensive		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Arcsinh(order count)	Entrants – short term	Entrants – 4 week window	Entrants – long term	Existing users – short term	Existing users - 2 month	Existing users – long term
Post × Digital	-0.054*	-0.049**	-0.003	-0.067***	0.018	0.104**
	(-2.49)	(-4.12)	(-0.17)	(-6.84)	(2.13)	(3.89)
Payment Mode FE (Digital dummy)	Yes	Yes	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9920	4960	9920	9952	9952	9952
R^2	0.953	0.955	0.949	0.970	0.967	0.968
Mean DV	2.965	2.973	2.979	4.033	4.008	4.019

Subsample analyses on extensive margin and intensive margin, respectively. The users for Columns 1 to 3 (entrants) is restricted to the observations for the first week each user uses the platform. Existing users in Columns 4 to 6 are defined as those ordering at least once per week from week 17 to week 20 (and before May 18th). We compare the month before the breach with the first month (Columns 1 and 4) and the third month ((Columns 3 and 6) after the breach respectively. The sample is restricted to a shorter time window (-2 weeks, 2 weeks) in Column 2. In Column 5, we compare the month before the breach with the second month after the breach. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table B.6 for Table A.4 Inattention: DDD using Google Trends

DV: Arcsinh(order count)	
Post × Digital × Google search	0.001
	(1.34)
Post × Digital	-0.178**
	(-4.24)
Digital × Google search	0.003
	(1.48)
Post × Google search	-0.001*
	(-2.78)
Digital	-0.036
	(-0.25)
Week FE zone FE, clustered at city level	Yes
Observations	10224
R^2	0.954
Mean DV	4.670

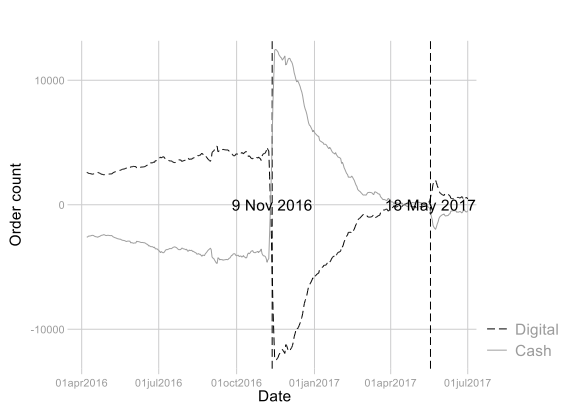
Subzone-payment mode-weekly level regression exploring how the response of payment modes to the data breach of the food delivery platform is related to Google search index. We compare the month before the breach with the first month after the breach. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

C. Appendix C: Poisson Regressions

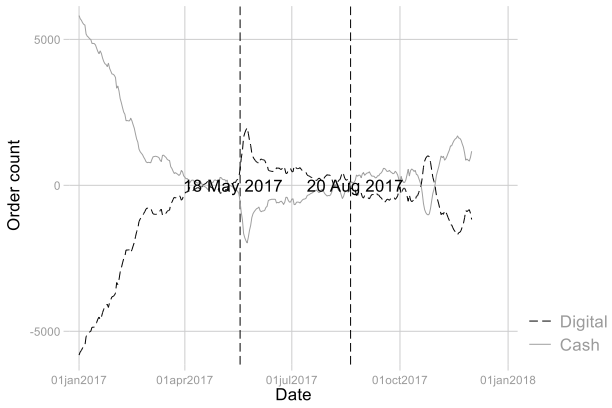
In this section, we provide robustness tests in which we use a count variable estimator. Another issue with the log-linear model is that in the presence of heteroskedasticity, the parameters of log-linear models estimated by OLS are inconsistent (Santos Silva and Tenreyro, 2006). In this context, using robust standard errors to mitigate concerns about heteroskedasticity will lead to incorrect inference because OLS estimates are not consistent in the first place. Using the alternative count variable estimator can circumvent the issue. Between two common count variable estimators – the Poisson regression and the negative binomial regression, we believe the Poisson model is more appropriate for our analysis for two reasons: (1) the Poisson regression estimator remains consistent even if the data are not Poisson-distributed as long as the conditional mean of the dependent variable is specified correctly (Gourieroux, Monfort, and Trognon, 1984) while the negative binomial regression requires strong distributional assumption of overdispersion. (2) Unlike most non-linear models, the Poisson model allows for separable fixed effects. We use the ppmlhdfe package for Stata (Correia, Guimarães, and Zylkin, 2020) to implement the Poisson Pseudo Maximum Likelihood (PPML) estimator with high-dimensional fixed effects. We apply the Poission model to all the previous regressions with 0 observations in the dependent variables $\ln(1+y)$ and present the estimates in the tables below. We include the original table (figure) numbers in the table (figure) titles for cross reference. Overall, the Poisson estimates are qualitatively similar to our OLS estimates of the log-linear and arcsinh-linear models.

Figure C.1 for Figure A.1 Demonetization

Panel A Long term



Panel B Short term



The figure shows the daily time series of the total number of orders paid by digital payments and cash payments respectively. The times series are seasonally adjusted by the method described in A.1.1 except that Poisson regression is used in the adjustment. 2 out of 1282 observations has 0 value for the order count.

Table C.1 for Table 1 Panel A Food Delivery Platform, Weekly

Dep. Var.: Order count	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.089***	-0.020**
	(-15.76)	(-2.28)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	9470	9656
R^2	0.962	0.963
Mean DV	278.997	277.755

Subzone-payment mode-weekly level Poisson regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table C.2 for Table 1 Panel B Food Delivery Platform, Sales

Dep. Var.: Sales	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.084***	0.015**
	(-10.10)	(2.42)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	9470	9624
Pseudo-R^2	0.976	0.977
Mean DV	113772	112636

Subzone-payment mode-weekly level Poisson regressions estimating the response of payment modes to the data breach of the food delivery platform. We use the two sample periods respectively and use sales as the dependent variable. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table C.3 for Table A.1 Food Delivery Platform, Daily

Dep. Var.: order count	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.069***	-0.018
	(-7.50)	(-1.41)
Payment Mode FE(Digital dummy)	Yes	Yes
Date × Subzone FE	Yes	Yes
Observations	69490	69974
Pseudo-R^2	0.913	0.915
Mean DV	42.414	42.686

Subzone-payment mode-daily level Poisson regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table C.4 for Table A.2 Panel A User level Analyses: Change in Payment Modes for the Food Delivery Platform Sample – Order Count

Dep. Var. Order count	(1) Post: the first month	(2) Post: the third month
Post × Digital	0.013	0.192***
	(0.30)	(9.06)
Payment Mode FE	Yes	Yes
User-Week FE	Yes	Yes
Observations	2587774	2353214
Pseudo-R^2	0.154	0.160
Mean DV	0.859	0.863

User-payment mode-weekly level Poisson regressions estimating the response of order placing and payment modes to the data breach. The sample is restricted to users who ordered at least once during the month right before the breach. Column (1) uses the short-term period and Column (2) uses the long-term period. Both sample periods are defined in Section A.2.2. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table C.5 for Table A.3 Extensive Margin and Intensive Margin

	Extensive			Intensive		
Dep. Var.: Order count	(1) Entrants – short term	(2) Entrants – 4 week window	(3) Entrants – long term	(4) Existing users – short term	(5) Existing users – 2 month	(6) Existing users – long term
Post × Digital	-0.062***	-0.055**	-0.037**	-0.064***	0.008	0.057***
	(-3.56)	(-2.02)	(-2.19)	(-11.02)	(0.44)	(3.44)
Payment Mode FE (Digital dummy)	Yes	Yes	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9136	4548	9268	9250	9332	9380
Pseudo R^2	0.871	0.874	0.865	0.961	0.961	0.962
Mean DV	31.412	32.246	29.951	157.451	150.899	149.425

Subsample analyses on extensive margin and intensive margin, respectively. The users for Columns 1 to 3 (entrants) is restricted to the observations for the first week each user uses the platform. Existing users in Columns 4 to 6 are defined as those ordering at least once per week from week 17 to week 20 (and before May 18th). We compare the month before the breach with the first month (Columns 1 and 4) and the third month (Columns 3 and 6) after the breach respectively. The sample is restricted to a shorter time window (-2 weeks, 2 weeks) in Column 2. In Column 5, we compare the month before the breach with the second month after the breach.

.

Table C.6 for Table A.4 Inattention: DDD using Google trends

DV: Order count	
Post × Digital × Google search	0.000
	(0.12)
Post × Digital	-0.114**
	(-2.38)
Digital × Google search	0.003*
	(1.88)
Post × Google search	0.000
	(0.30)
Digital	0.239*
	(1.77)
Week FE zone FE, clustered at week and city level	Yes
Observations	9776
Pseudo R^2	0.961
Mean DV	270.264

Subzone-payment mode-weekly level Poisson regression exploring how the response of payment modes to the data breach of the food delivery platform is related to Google search index. We compare the month before the breach with the first month after the breach. The mean dependent variable is reported at the bottom. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

D. Appendix D: Clustering of Standard Errors

In some of our baseline results, the standard errors are clustered at the city level, which leads to the potential downward bias of cluster-robust standard errors since the number of cities is small. In the baseline results, we rely on Stata's default small-sample correction—a simple multiplicative adjustment $m = \frac{G}{G-1} \times \frac{N-1}{N-k}$ to the asymptotic variance estimator, where *G* is the number of clusters (Imbens and Kolesar, 2016Imbens and Kolesar, 2016).[12] This choice has the intuitive property that, in the limiting case of $G = N$, that is, the "robust" case, the expression reduces to the classical $N/(N-k)$ correction factor. We use Stata's default degree of freedom adjustment in the significance testing: Stata uses the *t*(*G*-1) distribution to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters. As the *t*-distribution has heavier tails than the normal distribution, the critical value for a given level of significance is higher than the critical value under normality. As the number of degrees of freedom increases, the *t*-distribution approaches the standard normal distribution, the divergence of the critical values is substantial when we have a small number of clusters. For instance, for the analyses in Table 1, we have *G* = 4 clusters so the distribution used in significant testing is *t*(3) distribution. The critical values for significance at 5% and 10% level (two-sided) under the *t*(3) distribution are 3.182 and 5.841, respectively. As a comparison, the critical values for significance at 5% and 10% level (two-sided) under normality are 1.960 and 2.576, respectively.

Second, we also use the wild cluster bootstrapping method to correct for the small number of clusters (Cameron, Gelbach, and Miller, 2008; Cameron and Miller, 2015). In the Stata implementation, we use the "boottest" command (Roodman et al., 2019Roodman et al., 2019). A popular bootstrap procedure in applied literature is to use the standard deviation of the bootstrapped coefficient as the bootstrap estimate of standard error and then use the standard error estimate in a typical Wald test. Cameron, Gelbach, and Miller (2008) Cameron, Gelbach, and Miller (2008) caution that this "bootstrap-se" procedure may perform worse with few clusters than an alternative "bootstrap-t" procedure that computes the Wald statistic for each bootstrap replication and then uses the distribution of the bootstrap Wald statistics to calculate p-values both theoretically and through Monte Carlo simulations. Given that our issue at hand is precisely the small number of clusters, we opt for the "bootstrap-t" procedure and report the p-values, as opposed to the standard errors, from the wild cluster bootstraps at the bottom of each column in the baseline results with standard errors clustered at the city level (see the "P value for wild-t" row in related tables). We set the seed to 10000 to ensure replicability. The results are generally consistent.

In this section, we consider doubly clustering of the standard errors by city and by week to allow the error terms to be correlated over time. We apply the two-way clustering to the regressions with standard errors clustered at the city level and report the results below. We include the original table (figure) numbers in the table (figure) titles for cross reference. Note that the inverse hyperbolic sine transformation of the dependent variable is retained.

[12] Please also see the technical note in Stata User's Guide (https://www.stata.com/manuals/u.pdf) Section 20.22.2 *Correlated errors: Cluster–robust standard errors.*

We note that the following limitation of two-way clustering. The consistency of the clustered standard errors requires that the number of clusters goes to infinity. It is well-known that the cluster-robust standard errors are downwards biased when the number of clusters is small. The asymptotic assumption does become more binding with two-way or multiway clustering. For instance, in the case of two-way clustering it is assumed that min(G, H) goes to infinity, where there are G clusters in dimension 1 and H clusters in dimension 2. As a result, the finite-sample properties of multiway clustering further deteriorate. Cameron, Gelbach, and Miller (2011, Table 1) Cameron, Gelbach, and Miller (2011, Table 1) demonstrate through a Monte Carlo analysis that in a setting with ten clusters each in two dimensions, the one-way and two-way clustering methods over-reject even more than when i.i.d. errors are assumed. To see whether this issue is pertinent in our setting, we also report the t-statistics (or z-statistics for logit regressions) for specification without clustering, which are appropriate under i.i.d. errors. We note that in some of the regressions, the two-way clustered standard errors are even smaller than the OLS standard errors (namely, the absolute value of t-statistics for the two-way clustering specifications is bigger than those of specifications without clustering), suggesting of a downward bias.

Table D.1 for Table 1 Panel A Food Delivery Platform – Weekly

	(1)	(2)
Dep. Var.: Arcsinh(order count)	Post: the first month	Post: the third month
Post × Digital	-0.093**	0.024
	(-5.57)	(0.99)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	10224	10224
R^2	0.974	0.973
OLS t-stat for Post × Digital	-6.32	0.75
Mean DV	4.670	4.758

Subzone-payment mode-weekly (daily) level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. We apply two-way clustering at the city and week level to the standard errors. The *t*-statistics are reported in parentheses. The mean dependent variable is reported at the bottom. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.2 for Table 1 Panel B Food Delivery Platform – Sales

	(1)	(2)
Dep. Var.: Arcsinh(sales)	Post: the first month	Post: the third month
Post × Digital	-0.102***	0.098*
	(-6.33)	(3.02)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	10224	10224
R^2	0.949	0.948
OLS t-stat for Post × Digital	-3.51	1.46
Mean DV	10.084	10.281

Subzone-payment mode-weekly level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. We use the two sample periods respectively and use sales as the dependent variable. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where G is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.3 for Table 3 Panel B Quits and Payment Switches

Dep. Var.	Quit LOGIT Coef.		Δdigital_Ratio OLS Coef.		Δarcsinh(Digital payments) OLS Coef.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digital_ratio	-0.710***	-0.484***	-0.089***	-0.097***	-1.162***	-1.001***
	(-13.52)	(-11.67)	(-5.86)	(-6.16)	(-23.29)	(-34.19)
Quantity		-0.014		0.002		0.003
		(-0.81)		(1.06)		(1.07)
Account Age		-0.028***		0.002***		-0.027***
		(-6.92)		(7.57)		(-20.06)
Frequency		-0.378***		0.005		-0.182***
		(-7.52)		(2.09)		(-8.89)
Unit price		-0.003***		0.000**		0.000
		(-2.97)		(4.61)		(1.96)
Vegetarian index		-0.015		0.006		0.076
		(-0.67)		(0.63)		(1.23)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28930	28930	27117	27117	27117	27117
R^2			0.013	0.024	0.137	0.372
Pseudo R^2	0.014	0.063				
z-stat w/o clustering for Digital _ratio	-14.15	-18.43				
t-stat w/o clustering for Digital_ratio			-6.08	-6.63	-53.35	-40.83
Mean DV	0.179	0.179	0.014	0.014	-1.641	-1.641

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the change in the inverse hyperbolic sine transformation of the digital payment. The *t*-statistics are reported in parentheses. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where G is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.4 for Table A.1 Food Delivery Platform – Daily

Dep. Var.: Arcsinh(order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.068**	0.019
	(-4.68)	(1.23)
Payment Mode FE(Digital dummy)	Yes	Yes
Date × Subzone FE	Yes	Yes
Observations	79112	77836
R^2	0.962	0.961
t-stat w/o clustering for Post × Digital	-6.17	1.40
Mean DV	2.911	2.985

Subzone-payment mode-daily level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively. We apply two-way clustering at the city and week level to the standard errors. The *t*-statistics are reported in parentheses. The mean dependent variable is reported at the bottom. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.5 for Table A.3 Extensive Margin and Intensive Margin

	Extensive			Intensive		
Dep. Var.: Arcsinh(order count)	(1) Entrants – short term	(2) Entrants – 4 week window	(3) Entrants – long term	(4) Existing users – short term	(5) Existing users – 2 month	(6) Existing users – long term
Post × Digital	-0.054***	-0.049**	-0.003	-0.067**	0.018	0.104***
	(-10.10)	(-4.62)	(-0.23)	(-5.49)	(0.98)	(11.24)
Payment Mode FE (Digital dummy)	Yes	Yes	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9920	4960	9920	9952	9952	9952
Pseudo R^2	0.953	0.955	0.949	0.970	0.967	0.968
t-stat w/o clustering for Post × Digital	-2.76	-4.79	-0.35	-7.23	2.39	4.19
Mean DV	2.965	2.973	2.979	4.033	4.008	4.019

Subsample analyses on extensive margin and intensive margin, respectively. The users for Columns 1 to 3 (entrants) is restricted to the observations for the first week each user uses the platform. Existing users in Columns 4 to 6 are defined as those ordering at least once per week from week 17 to week 20 (and before May 18th). We compare the month before the breach with the first month (Columns 1 and 4) and the third month ((Columns 3 and 6) after the breach respectively. The sample is restricted to a shorter time window (-2 weeks, 2 weeks) in Column 2. In Column 5, we compare the month before the breach with the second month after the breach. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where G is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.6 for Table A.4 Inattention: DDD using Google Trends

DV: Arcsinh (order count)	
Post × Digital × Google search	0.001
	(1.62)
Post × Digital	-0.186**
	(-3.91)
Digital × Google search	0.003
	(1.92)
Post × Google search	-0.001
	(-2.22)
Digital	-0.028
	(-0.23)
Week FE zone FE, clustered at week and city level	Yes
Observations	10224
R^2	0.954
t-stat w/o clustering for Post × Digital × Google search	1.20
Mean DV	4.670

Subzone-payment mode-weekly level regression exploring how the response of payment modes to the data breach of the food delivery platform is related to Google search index. We compare the month before the breach with the first month after the breach. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.7 for Table A.6: Changes in Payment Modes for Grocery Store

	(1)	(2)
Dep. Var.: Arcsinh(order count)	Weekly	Daily
Post × Digital	0.060***	0.057***
	(4.99)	(6.86)
Payment Mode FE(Digital dummy)	Yes	Yes
Week × City FE	Yes	
Date × City FE		Yes
Observations	96	732
R^2	0.989	0.977
t-stat w/o clustering for Post × Digital	1.46	1.11
Mean DV	7.075	5.110

City-payment mode-weekly (daily) level regressions estimating the response of payment modes to the data breach. We compare the month before and after the Facebook data breach. The dependent variable is the natural logarithm of the number of orders. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table D.8 for Table A.7 Panel B Grocery Store: Quits and Payment Switches

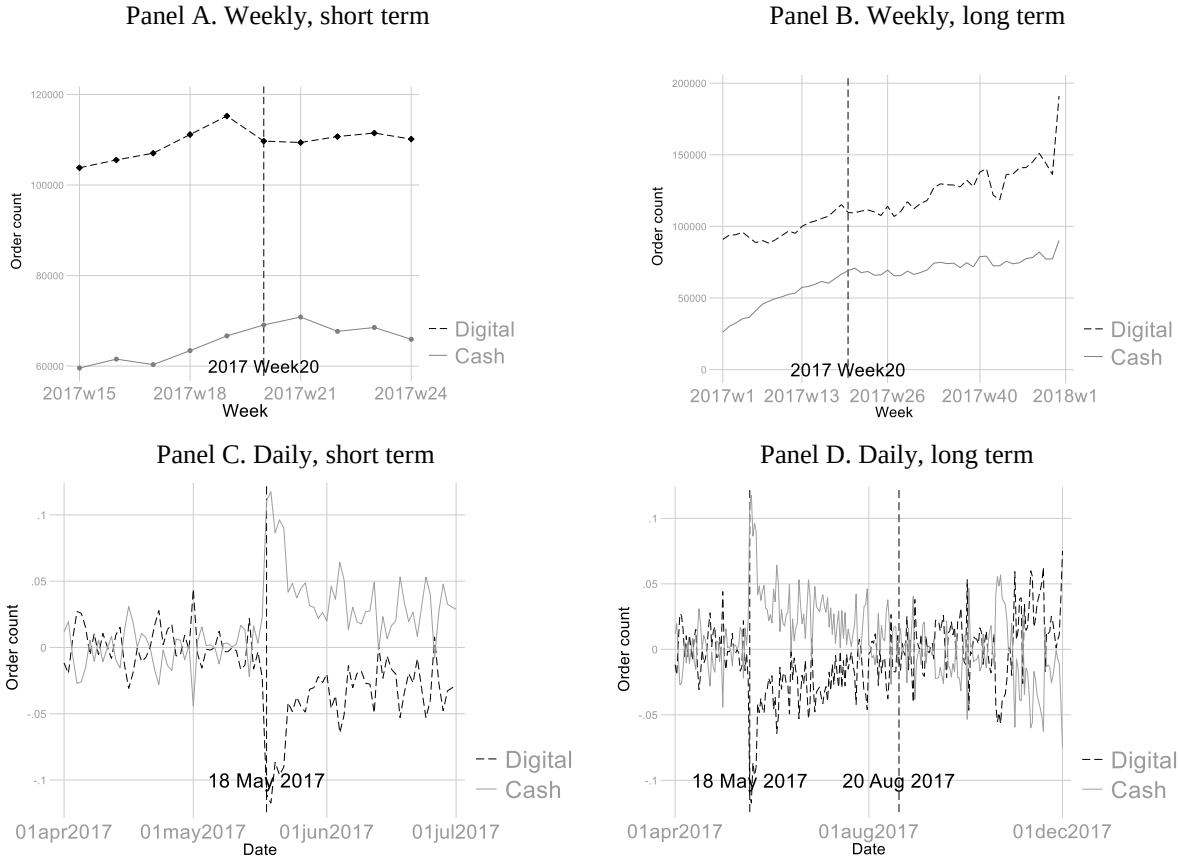
Dep. Var.	Quit		Δdigital_Ratio		Δarcsinh(digital payments)	
	LOGIT Coeff.		OLS Coeff.		OLS Coeff.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digital_ratio	-1.474***	-0.848	-0.206***	-0.200***	-1.092***	-1.113***
	(-4.87)	(-1.53)	(-5.34)	(-4.60)	(-15.87)	(-11.82)
Quantity		-0.061**		0.002*		0.001
		(-2.46)		(2.23)		(0.39)
Account Age				0.001***		-0.016***
				(7.42)		(-40.29)
Frequency		-2.370***		-0.002		-0.077***
		(-10.33)		(-0.48)		(-6.56)
Price		0.001***		-0.000		0.000**
		(2.70)		(-1.37)		(3.90)
City FE			Yes	Yes	Yes	Yes
Observations	815	815	814	814	814	814
R^2			0.080	0.084	0.205	0.281
Pseudo R^2	0.012	0.151				
z-stat w/o clustering for Digital_ratio	-2.80	-1.30				
t-stat w/o clustering for Digital_ratio			-4.46	-4.26	-6.69	-6.96
Mean DV	0.002	0.002	0.114	0.114	-3.203	-3.203

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample is for the online grocery store. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the change in the inverse hyperbolic sine transformation of the digital payment. Robust standard errors are doubly clustered at the city level and at the week level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

E. Appendix E: Overlapped Cities

In this section, we restrict the samples to two cities that appear in all datasets, Bangalore and Mumbai, and repeat all the analyses. Data for Figure A.7 are not available at the city level, and the analysis is not repeated. Note the sample size is largely reduced after we restrict the samples to the two cities, so some samples may not have sufficient statistical power, and the results should be interpreted cautiously. The results are listed below. We include the original table (figure) numbers in the table (figure) titles for cross reference. Note that the inverse hyperbolic sine transformation of the dependent variable is retained in the regressions for all the tables in this section. For seasonality adjustment of the daily samples in Figure E.1 and Figure E.3, we apply inverse hyperbolic sine transformation to the dependent variable since some observations have zero values for the dependent variable.

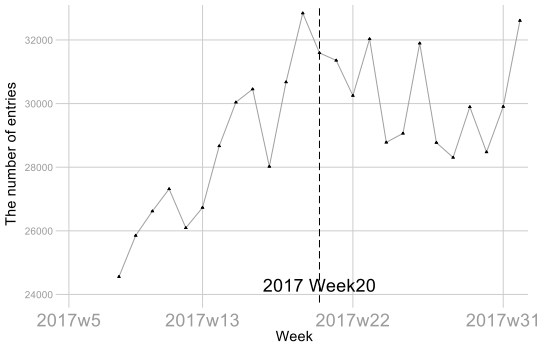
Figure E.1 for Figure 1 Changes in Payment Modes – Order Count



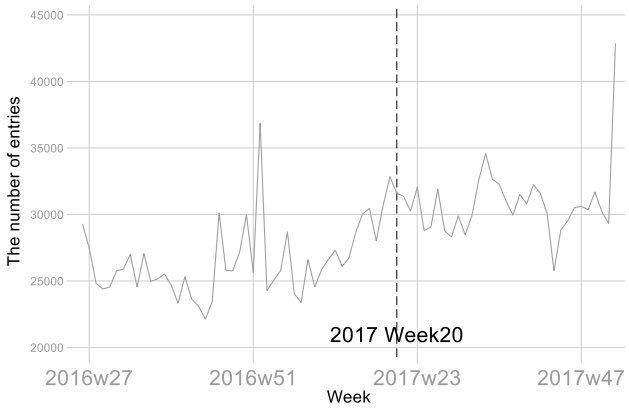
The figure shows the time series of the number of orders paid by digital and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample. Panels C and D are for the seasonally adjusted daily sample mentioned in Section A.1.1. In the daily sample, 1 out of 5 observations has zero value for the dependent variable, and we apply inverse hyperbolic sine transformation on the dependent variable.

Figure E.2 for Figure 2 Consumer Entries and Quits

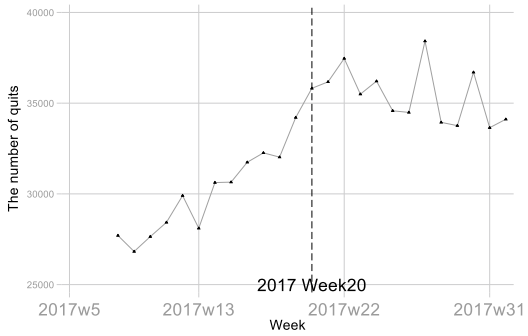
Panel A. Consumer Entries, short term



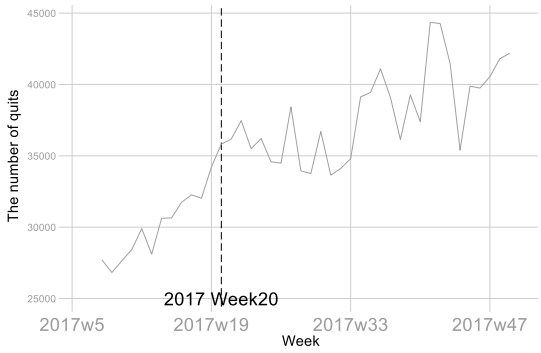
Panel B. Consumer Entries, long term



Panel C. Consumer Quits, short term



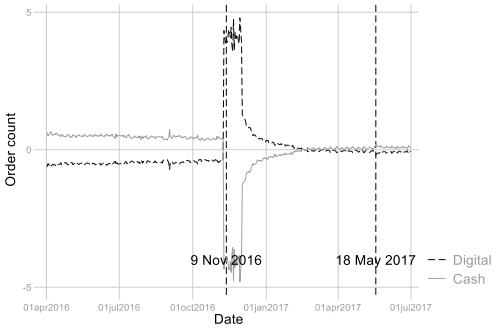
Panel D. Consumer Quits, long term



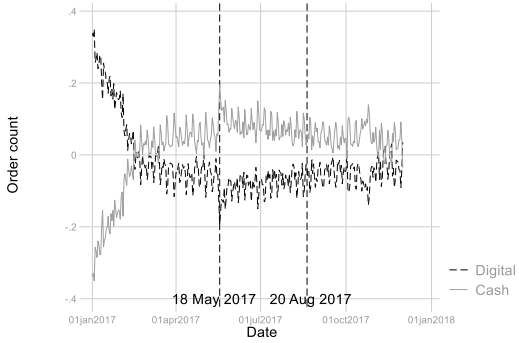
The figure shows the time series of consumer entries and quits. Panels A and B are for entries that are defined as the first appearance of the user in the sample. Panels C and D are for quits. The week of a quit is defined as the first week of four-week inactivity.

Figure E.3 for Figure A.1 Demonetization

Panel A Long term

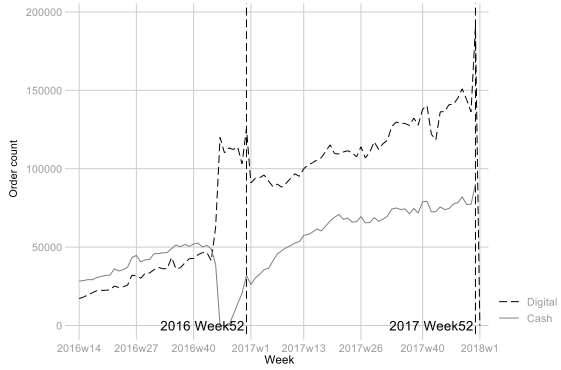


Panel B Short term



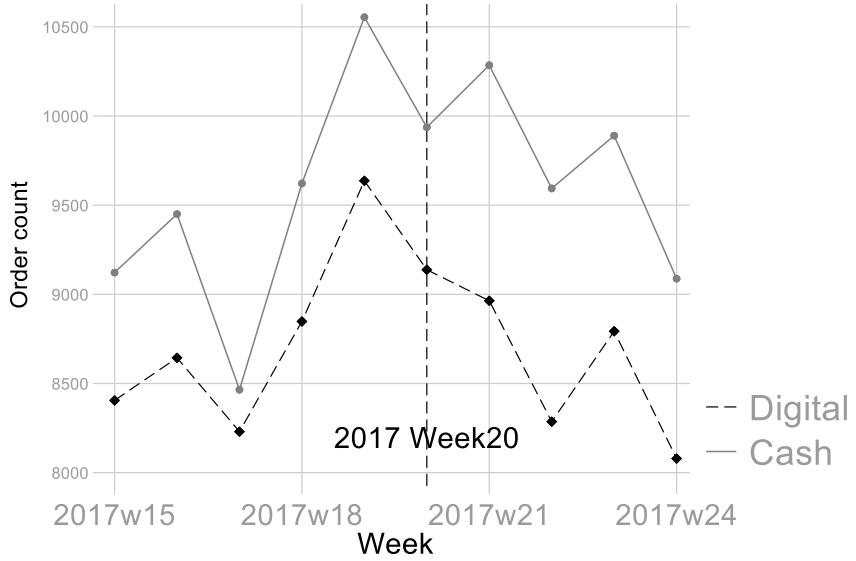
The figure shows the daily time series of the total number of orders paid by digital payments and cash payments respectively. The times series are seasonally adjusted by the method described in A.1.1. 3 out 1282 observations have zero values for the dependent variable, and we apply inverse hyperbolic sine transformation on the dependent variable.

Figure E.4 for Figure A.2 Year-end spikes



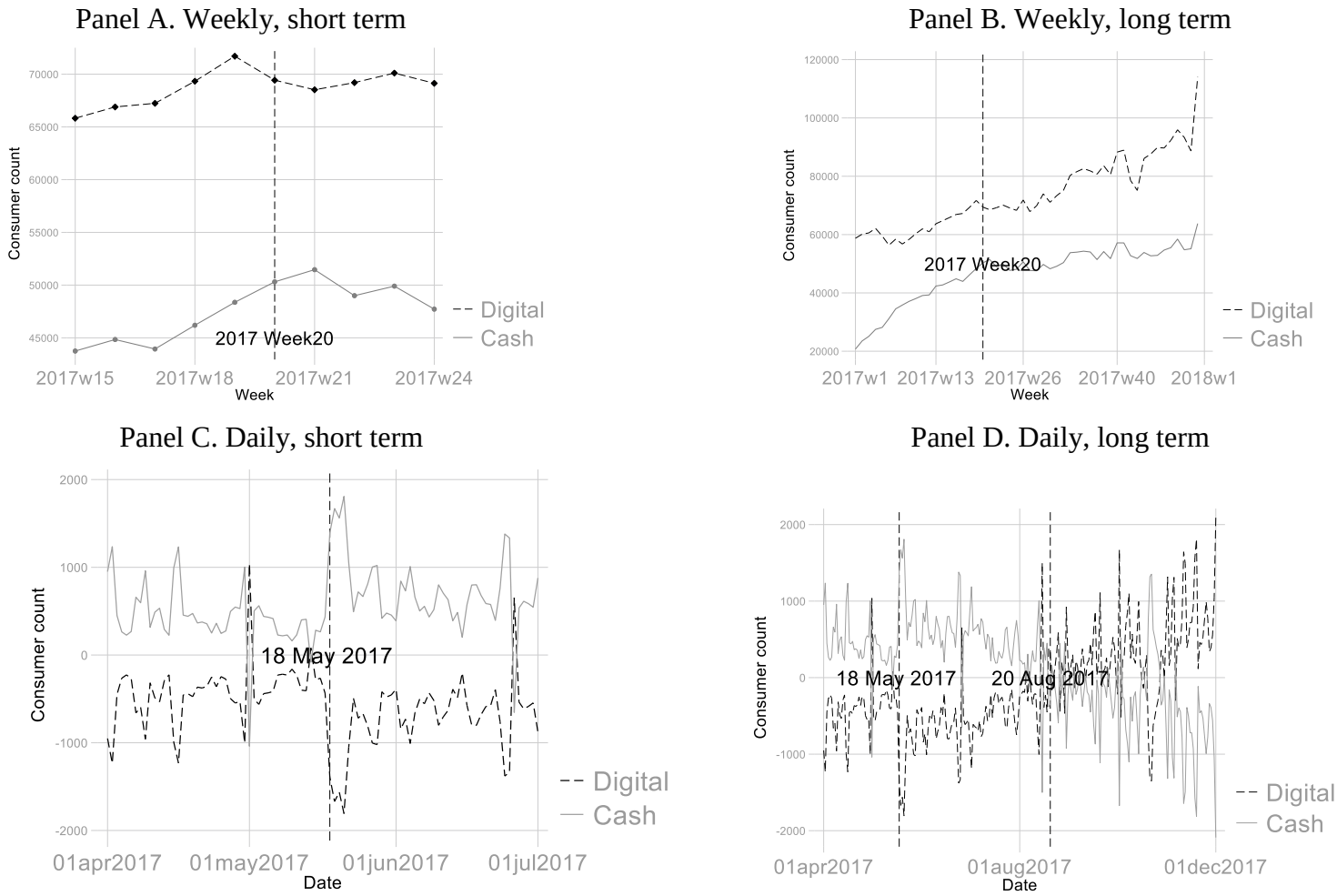
The figure shows the weekly time series of the total number of orders paid by digital payments and cash payments respectively.

Figure E.5 for Figure A.4 Entrants: Cash and Digital Payments



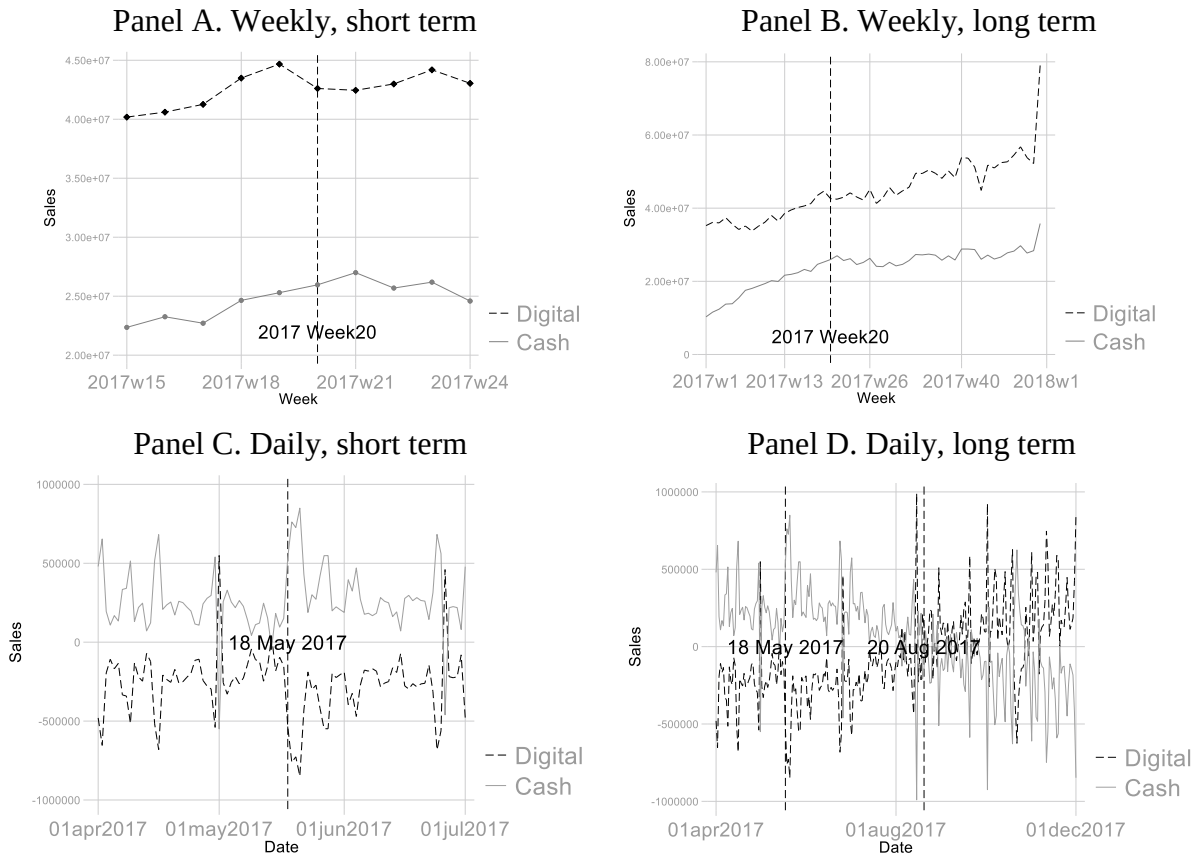
The figure shows the weekly time series of order counts for cash and digital payments by entrants (i.e. during the week they appear in the platform for the first time.

Figure E.6 for Figure A.5 Changes in Payment Modes - Consumer Count



The figure shows the time series of the number of consumers who use digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample, and Panels C and D are for the daily sample. The daily sample is seasonally adjusted as mentioned in Section A.1.1.

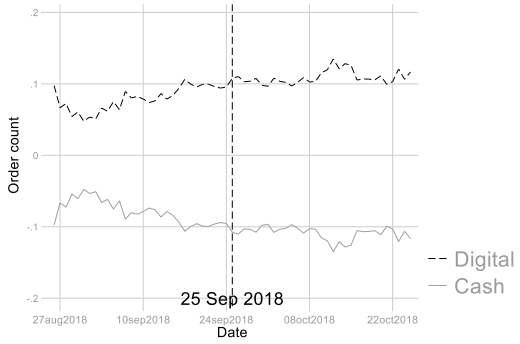
Figure E.7 for Figure A.6 Changes in Payment Modes -Sales



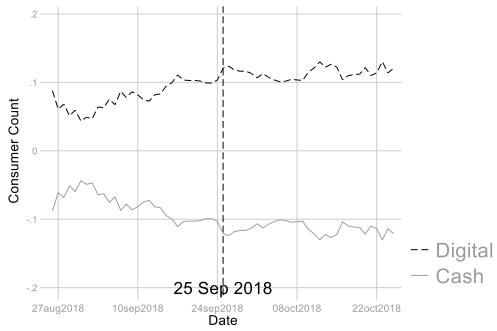
The figure shows the time series of the rupee volume of sales for digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. Panels A and B are for the weekly sample, and Panels C and D are for the daily sample. The daily sample is seasonally adjusted as mentioned in Section A.1.1. The calculation of sales is introduced in Section A.2.5.

Figure E.8 for Figure A.8 Grocery Store: Changes in Payment Modes

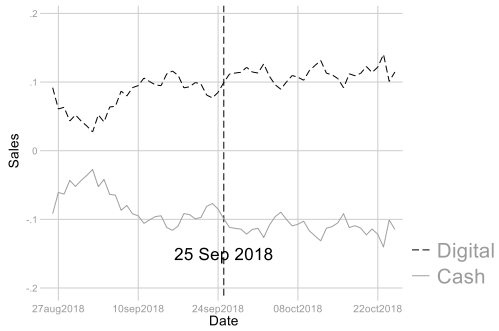
Panel A. Order count



Panel B. User count



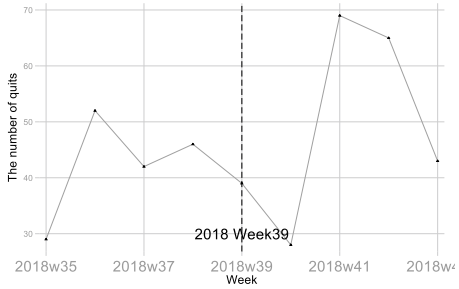
Panel C. Sales



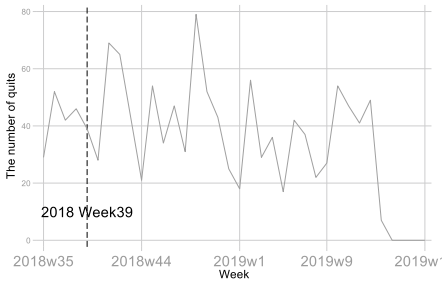
The figure shows the daily time series of the user count, order count, and the rupee volume of sales for digital payments and cash payments respectively. Dashed lines represent digital payments, and solid lines represent cash payments. The time series is for the sample of the online grocery store and is seasonally adjusted as mentioned in Section A.4.2.

Figure E.9 for Figure A.9 Grocery Store: User Quits

Panel A. Short Term

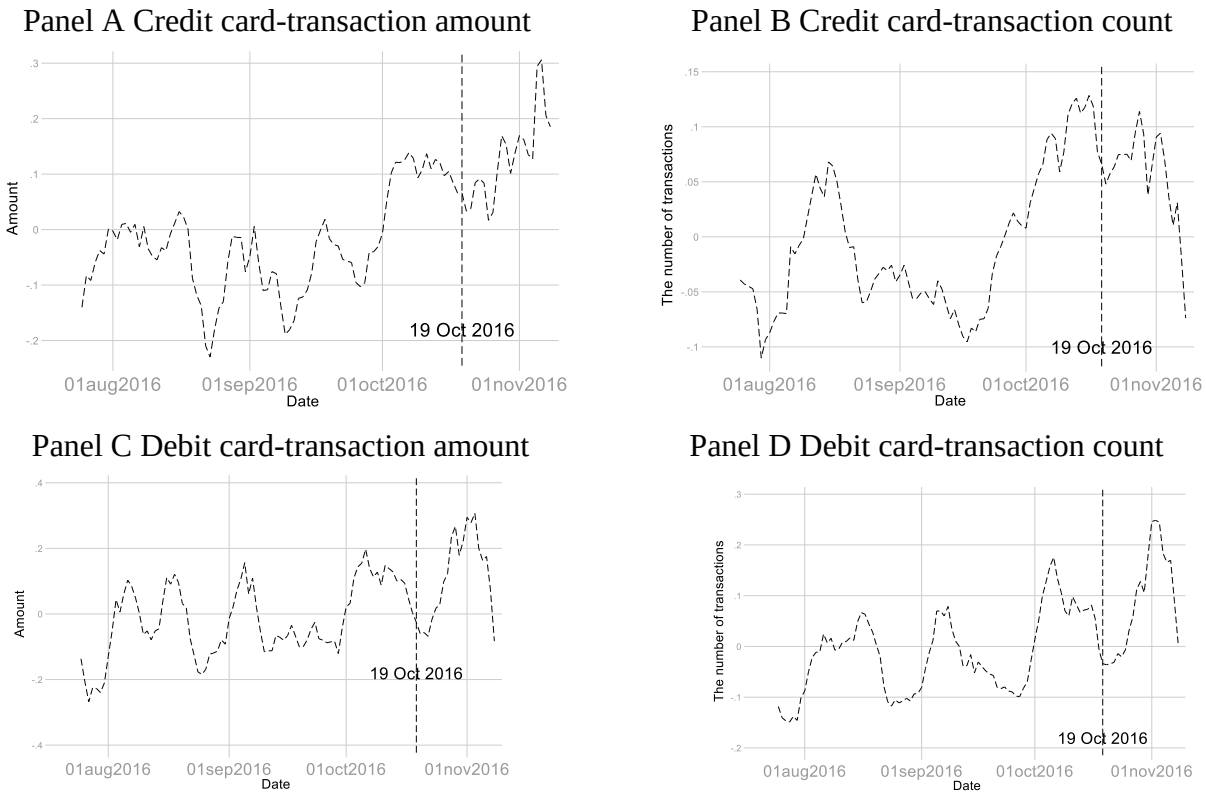


Panel B. Long Term



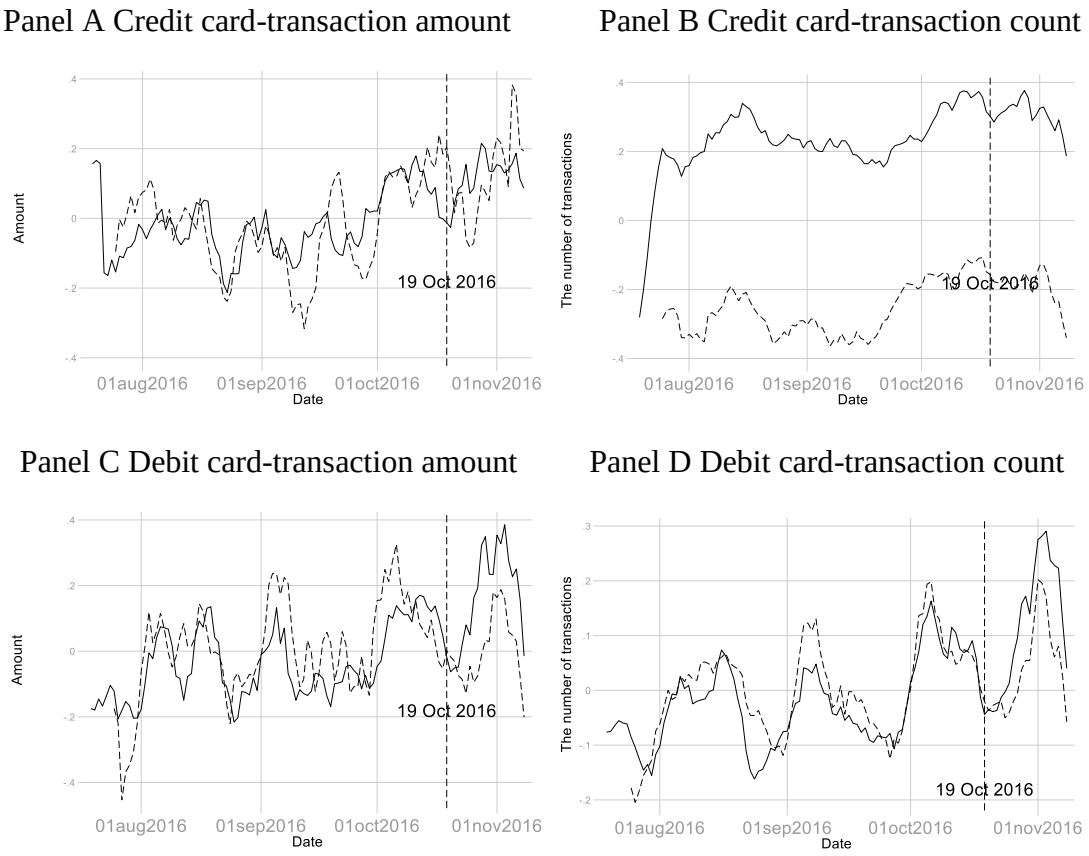
The figure shows the weekly time series of consumer quits. The sample is for the online grocery store. The week of a quit is defined as the first week of a four-week inactivity.

Figure E.10 for Figure A.10 Transaction Change of Bank Users: the Data Breach in Indian Banking System



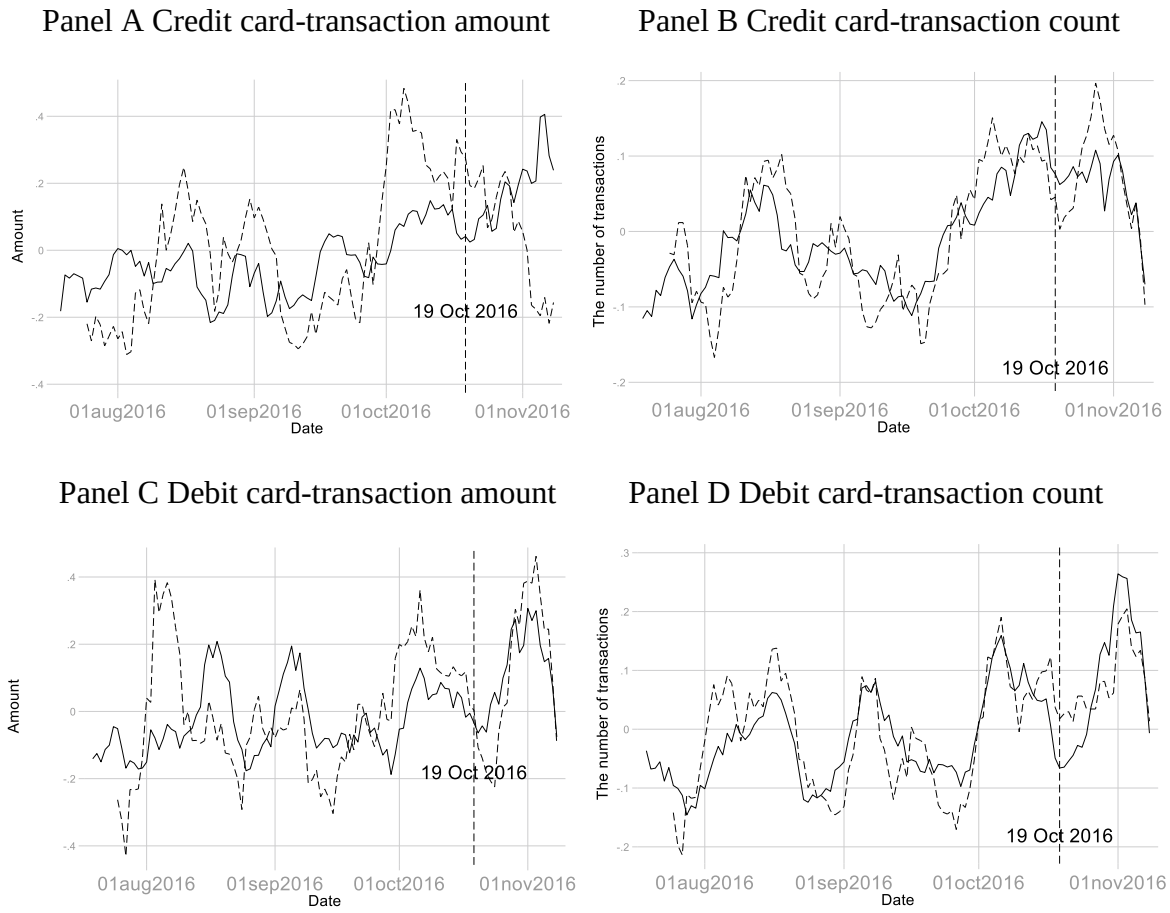
The figure shows the effects of the bank data breach on the transactions of the bank users. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Figure E.11 for Figure A.11 Transaction Change of Bank Users: the Data Breach in Indian Banking System – by Account Balance



The figure shows the effects of the bank data breach on the transactions of the high–wealth and low-wealth bank users defined in Section A.5.2. The solid line represents high-wealth users, and the dash line represents the low-wealth users. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Figure E.12 for Figure A.12 Transaction Change of Bank Users: the Data Breach in Indian Banking System – By Credit Limit



The figure shows the effects of the bank data breach on the transactions of the bank users with high and low credit limit defined in Section A.5.2. The solid line represents users with high credit limit, and the dash line represents the users with low credit limit. It plots the daily time series of the transaction count and the rupee volume of transaction. The time series is for the sample of the bank customers and is seasonally adjusted as mentioned in Section A.5.2.

Table E.1 for Table 1 Changes in Payment Modes

Panel A Weekly

Dep. var.: Arcsinh(order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.055**	0.098*
	(-12.79)	(6.84)
Payment Mode FE (Digital dummy)	Yes	Yes
Week × Subzone FE	Yes	Yes
Observations	3360	3360
R^2	0.972	0.976
Mean DV	5.553	5.586

Panel B Alternative Dependent Variables

Dep. var.:	(1) Arcsinh(sales)	(2) Arcsinh (order price)	(3) Arcsinh (sales)	(4) Arcsinh (order price)
	Post: the first month		Post: the third month	
Post × Digital	-0.043*	0.008	0.251*	0.040*
	(-7.34)	(0.88)	(10.02)	(7.22)
Payment Mode FE(Digital dummy)	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes
Observations	3360	3158	3360	3200
R^2	0.947	0.785	0.953	0.808
Mean DV	11.212	6.651	11.294	6.635

Subzone-payment mode-weekly level regressions estimating the response of payment modes to the data breach of the food delivery platform. In Panel A, we compare the month before the breach with the first and the third month after the breach respectively, and the dependent variable is the natural logarithm of the number of orders. In Panel B, we use the two sample periods respectively and replace the dependent variable with sales and the price of orders, respectively. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.2 for Table 2 Data Breach and Customer Entries

Variable	Before	After	Diff/P-value
Price	1444.408	1456.3	11.8911
			[0.5604]
Unit Price	187.384	191.149	3.7651***
			[0.0000]
Quantity	2.943	2.929	-0.0136
			[0.4159]
Vegetarian Index	0.357	0.353	-0.0044*
			[0.0926]
Observations	67369	69164	136533

This table shows the t-tests comparing the means of the characteristics of the first order by the customers who entered one month before and after the data breach.

Table E.3 for Table 3 Quits and Payment Switches

Panel A Summary Statistics

Variable	Obs	Mean	Std. Dev.
Quit	15,201	0.17	0.38
Digital_ratio	15,201	0.67	0.29
Quantity	15,201	2.58	1.11
Account Age	15,201	39.16	17.61
Frequency	15,201	1.66	1.25
Unit price	15,201	160.03	46.10
Vegetarian Index	15,201	0.20	0.16
ΔDigital_Ratio	14,325	0.02	0.25

Panel B Regression Results

Dep. var.	Quit		ΔDigital_Ratio		ΔArcsinh(Digital payments)	
	LOGIT Coef.		OLS Coef.		OLS Coef.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digtal_ratio	-0.690***	-0.496***	-0.107	-0.116	-1.157**	-1.011**
	(-38.55)	(-565.42)	(-4.01)	(-4.15)	(-49.04)	(-43.00)
Quantity		-0.002		0.003		0.004
		(-0.30)		(4.73)		(1.70)
Account Age		-0.026***		0.002		-0.027**
		(-12.08)		(4.78)		(-47.58)
Frequency		-0.383***		0.006		-0.165**
		(-302.98)		(5.84)		(-20.70)
Unit price		-0.002***		0.000		0.000
		(-4.56)		(3.60)		(1.28)
Vegetarian index		0.037		-0.004		0.026
		(0.39)		(-0.43)		(0.32)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15201	15201	14325	14325	14325	14325
R^2			0.019	0.032	0.126	0.368
Pseudo R^2	0.015	0.057				
Mean DV	0.170	0.170	0.020	0.020	-1.644	-1.644
Marginal Effects of	-0.096***	-0.066***				
Digital_Ratio	(-39.-07)	(-156.98)				

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. Panel A displays the descriptive statistics of the sample, and Panel B displays the estimates. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the natural logarithm of the digital payment. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where G is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.4 for Table A.1 Food Delivery Platform, Daily

Dep. var.: Arcsinh(order count)	(1) Post: the first month	(2) Post: the third month
Post × Digital	-0.035*	0.050
	(-6.83)	(1.94)
Payment Mode FE(Digital dummy)	Yes	Yes
Date × Subzone FE	Yes	Yes
Observations	25916	25498
R^2	0.964	0.966
Mean DV	3.694	3.727

Subzone-payment mode-daily level regressions estimating the response of payment modes to the data breach of the food delivery platform. We compare the month before the breach with the first and the third month after the breach respectively, and the dependent variable is the natural logarithm of the number of orders. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.5 for Table A.2 Panel A User-level Analyses: Change in Payment Modes for the Food Delivery Platform Sample

Dep. var.	(1) Arcsinh(order count)	(2) Digital payment ratio	(3) Arcsinh(order count)-long term	(4) Digital payment ratio-long term
Post × Digital	-0.022***		-0.020***	
	(-4.16)		(-4.06)	
Post		-0.013***		0.011***
		(-9.63)		(6.35)
Payment Mode FE	Yes		Yes	
User-Week FE	Yes		Yes	
User FE		Yes		Yes
Observations	3833978	203737	3448342	181867
R^2	0.515	0.749	0.225	0.739
Mean DV	0.222	0.675	0.202	0.690

User-payment mode--weekly level regressions estimating the response of order placing and payment modes to the data breach. The sample is restricted to users who ordered at least once during the month right before the breach. Column (1) and (2) use the short-term period and Columns (3) and (4) use the long-term period. Both sample periods are defined in Section A.2.2. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.6 for Table A.3 Extensive Margin and Intensive Margin

	Extensive			Intensive		
Dep. var.: Arcsinh(order count)	(1) Entrants – short term	(2) Entrants – 4 week window	(3) Entrants – long term	(4) Existing users – short term	(5) Existing users - 2 month	(6) Existing users – long term
Post × Digital	-0.012	-0.043	0.045	-0.066	0.034	0.112
	(-0.53)	(-1.66)	(2.11)	(-2.09)	(4.38)	(2.39)
Payment Mode FE (Digital dummy)	Yes	Yes	Yes	Yes	Yes	Yes
Week × Subzone FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3296	1648	3296	3312	3312	3312
R^2	0.956	0.957	0.956	0.973	0.971	0.973
Mean DV	3.670	3.701	3.647	4.858	4.811	4.814

Subsample analyses on extensive margin and intensive margin, respectively. The users for Columns 1 to 3 (entrants) is restricted to the observations for the first week each user uses the platform. Existing users in Columns 4 to 6 are defined as those ordering at least once per week from week 17 to week 20 (and before May 18th). We compare the month before the breach with the first month (Columns 1 and 4) and the third month ((Columns 3 and 6) after the breach respectively. The sample is restricted to a shorter time window (-2 weeks, 72 weeks) in Column 2. In Column 5, we compare the month before the breach with the second month after the breach.

Table E.7 for Table A.4 Inattention: DDD using Google trends

DV: Arcsinh(order count)	
Post × Digital × Google search	-0.001
	(-3.75)
Post × Digital	0.035
	(3.92)
Digital × Google search	-0.007*
	(-7.21)
Post × Google search	0.000
	(0.16)
Digital	0.869**
	(18.29)
Week FE zone FE, clustered at city level	Yes
Observations	3360
R^2	0.965
Mean DV	5.553

Subzone-payment mode-weekly level regressions exploring how the response of payment modes to the data breach of the food delivery platform is related to Google search index. We compare the month before the breach with the first month after the breach, and the dependent variable is the natural logarithm of the number of orders. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.8 for Table A.6 Changes in Payment Modes for Grocery Store

Dep. var.: Arcsinh(order count)	(1) Weekly	(2) Daily
Post × Digital	0.080	0.069
	(2.23)	(1.47)
Payment Mode FE (Digital dummy)	Yes	Yes
Week × City FE	Yes	
Date × City FE		Yes
Observations	32	244
R^2	0.964	0.959
Mean DV	8.008	6.048

City-payment mode-weekly (daily) level regressions estimating the response of payment modes to the data breach. We compare the month before and after the Facebook data breach. The dependent variable is the natural logarithm of the number of orders. Robust standard errors are clustered at the city level. The *t*-statistics are reported in parentheses. For significance testing, the $t(G\text{-}1)$ distribution is used to obtain p-values for cluster-robust *t*-statistics where *G* is the number of clusters; *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table E.9 For Table A.7 Grocery Store: Quits and Payment Switches

Panel A Summary Statistics

Variable	Obs	Mean	Std. Dev.
Quit	536	0.004	0.06
Digital_ratio	536	0.772	0.23
Quantity	536	15.558	7.80
Account Age	536	141.077	7.35
Frequency	536	2.286	1.30
Unit price	536	1.077	0.10
ΔDigital_Ratio	535	0.105	0.18

Panel B Regression Results

Dep. var.	Quit		ΔDigital_ratio		ΔArcsinh(digital payments)	
	LOGIT Coeff.		OLS Coeff.		OLS Coeff.	
	(1)	(2)	(3)	(4)	(5)	(6)
Digital_ratio	-1.901***	-0.966	-0.214**	-0.210**	-0.981**	-0.989**
	(-2.68)	(-0.95)	(-16.62)	(-31.63)	(-18.09)	(-14.14)
Quantity		-0.064***		0.002		-0.004
		(-3.74)		(3.83)		(-0.65)
Account Age				0.001		-0.017***
				(1.56)		(-90.09)
Frequency		-2.436***		0.002		-0.079***
		(-56.77)		(0.91)		(-669.62)
Price		0.001**		-0.000		0.000
		(2.01)		(-2.00)		(0.48)
City FE			Yes	Yes	Yes	Yes
Observations	536	536	535	535	535	535
R^2			0.072	0.076	0.159	0.237
Pseudo R^2	0.021	0.158				
Mean DV	0.004	0.004	0.105	0.105	-3.210	-3.210
Marginal effects of Digital_ratio	-0.007	-0.004***				
	(-0.92)	(-3.41)				

User-level regressions estimating the quits and payment switches of consumers after the data breach event. The sample is for the online grocery store. The sample period is cross-sectional and is constructed based on the period between the month before and the month after the breach event. Panel A displays the descriptive statistics of the sample, and Panel B displays the estimates. In Columns 1 and 2, the dependent variable is an indicator that equals 1 if the consumer quit within four weeks after the breach. In Columns 3 and 4, the dependent variable is the change in the digital payment ratio. In Columns 5 and 6, the dependent variable is the change in the inverse hyperbolic sine transformation of the digital payment. The t-statistics are reported in parentheses. The mean dependent variable is reported at the bottom to assess marginal effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

References for the Internet Appendix

Ablon, Lillian, Paul Heaton, Diana Catherine Lavery, and Sasha Romanosky. 2016. "Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information." doi: 10.7249/RR1187.

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. 2020. "Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age." *Journal of Consumer Psychology* no. 30 (4):736-758.

Agarwal, Sumit, Pulak Ghosh, Jing Li, and Tianyue Ruan. 2019. "Digital payments induce over-spending: Evidence from the 2016 demonetization in India." *Working paper*.

Athey, Susan, Christian Catalini, and Catherine Tucker. 2017. "The digital privacy paradox: Small money, small costs, small talk." *NBER Working Paper No.w23488*.

Bellemare, Marc F., and Casey J. Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics* no. 82:50-61.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-differences Estimates?" *Quarterly Journal of Economics*.

Burlig, Fiona, Louis Preonas, and Matt Woerman. 2020. "Panel data and experimental design." *Journal of Development Economics* no. 144:102458. doi: 10.1016/j.jdeveco.2020.102458.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics* no. 90:414-427. doi: 10.1162/rest.90.3.414.

———. 2011. "Robust inference with multiway clustering." *Journal of Business and Economic Statistics* no. 29:238-249. doi: 10.1198/jbes.2010.07136.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resource* no. 50:317-372. doi: 10.3368/jhr.50.2.317.

Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra, and Abhinav Narayanan. 2020. "Cash and the economy: Evidence from India's demonetization." *The Quarterly Journal of Economics* no. 135 (1):57-103.

Cohen, J. 1988. Stafisfical power analysis for the behavioural sciences (2nd edn.). Hillside. NJ: Erlbaum.

Cohn, Jonathan B., Zack Liu, and Malcolm Wardlaw. 2021. "Count Data in Finance." *SSRN Electronic Journal*. doi: 10.2139/ssrn.3800339.

Correia, Sergio, Paulo Guimarães, and Tom Z. Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *Stata Journal* no. 20:95-115. doi: 10.1177/1536867X20909691.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G Rossi. 2019. "The promises and pitfalls of robo-advising." *The Review of Financial Studies* no. 32 (5):1983-2020.

D'Acunto, Francesco, Thomas Rauter, Christoph K Scheuch, and Michael Weber. 2020. "Perceived precautionary savings motives: Evidence from fintech." *NBER Working Paper*.

Duan, Naihua, Willard G Manning, Carl N Morris, and Joseph P Newhouse. 1983. "A comparison of alternative models for the demand for medical care." *J Journal of business economic statistics* no. 1 (2):115-126.

Frison, Lars, and Stuart J Pocock. 1992. "Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design." *Statistics in medicine* no. 11 (13):1685-1704.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo maximum likelihood methods: Applications to Poisson models." *Econometrica*:701-720.

Imbens, Guido W, and Michal Kolesar. 2016. "Robust standard errors in small samples: Some practical advice." *Review of Economics Statistics* no. 98 (4):701-712.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* no. 99:210-221. doi: 10.1016/j.jdeveco.2012.01.002.

N'guessan, Yapo Genevier, Allen Featherstone, Oluwarotimi Odeh, and Sreedhar Upendram. 2017. "Choice of the empirical definition of zero in the translog multiproduct cost functional form." *Applied Economics Letters* no. 24 (15):1112-1120.

Roodman, David, James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb. 2019. "Fast and wild: Bootstrap inference in Stata using boottest." *Stata Journal* no. 19:4-60. doi: 10.1177/1536867X19830877.

Santos Silva, J. M.C., and Silvana Tenreyro. 2006. "The log of gravity." *Review of Economics and Statistics* no. 88:641-658.