Online Appendix for "Shocks and Technology Adoption: Evidence from Electronic Payment Systems"

This version: January, 2023

A Institutional background

This section is organized in three parts. First, we describe the 2016 Demonetization in more detail, highlighting the features of this event that are the most relevant for our analysis. Second, we discuss the role of the government in the post-shock period. Third, we present a detailed discussion of the technology we are focusing on: mobile wallet electronic payment.

A.1 The economic impact of the Demonetization

The event The announcement of the Demonetization on November 8, 2016 voided about 86.4% of the total value of currency in circulation automatically. Even though the Indian population had until the end of the year to deposit the old notes in the banking sector, the voided bills could not be used immediately after the announcement. At the same time, the new notes were not available right away, as the central bank had not even finished printing all the necessary bills in November. Combining these two things together, India found itself with a shortage of currency in cash overnight.

Evidence of the scarcity of cash is abundant during this period. One manifestation is the disruption that characterized banks' operation during this period. In a survey of 214 households in 28 slums in the city of Mumbai, 88% of households reported waiting for more than 1 hour for ATM or bank services between 11/09/2016 and 11/18/2016. In the same survey, 25% of households reported waiting for more than 4 hours (Krishnan and Siegel, 2017). Another randomized survey conducted over nine districts in India by a mainstream newspaper, Economic Times, showed that the number of visits to either a bank or an ATM increased from an average of 5.8 in the month before Demonetization to 14.4 in the month after Demonetization.[81] This evidence confirms the presence of a large unmet demand for cash during the aftermath of the Demonetization.

For consumers, the generalized scarcity of cash was made worse by the constraints on cash withdrawal that were put in place by the government. In its initial press release, the RBI indicated that over the counter cash exchanges could not exceed Rs.4,000 per person per day, while withdrawals from accounts were capped at Rs.20,000 per week, and ATM withdrawals were capped at Rs.4,000 per card per day, for the days following the announcement. However, a wide set of exceptions were granted, including for fuel pumps, toll payments, government hospitals, and wedding expenditures.[82] Banerjee et al. (2018) discuss the uncertainty surrounding the withdrawal limits and exceptions, and argue that this uncertainty may have exacerbated the overall confusion during this transition period.

In sum, there are two features of the shock that are worth highlighting. First, the policy led to a large and extremely significant reduction in the availability of cash, which generated a constraint on households' ability to conduct transaction using cash. This claim is consistent with the evidence on the economic costs of the Demonetization. For instance, our own analyses highlight how households in areas more affected by the shock experienced a temporary reduction in consumption (Appendix D), consistent with the results of Chodorow-Reich et al. (2019). Second, the shock did not change the total wealth of households, but only the ability to utilize cash. In fact, the public could still deposit the notes, and access them using non-cash

[81] economictimes.indiatimes.com/news/politics-and-nation/how-delhi-lost-a-working-day-to-demonetisation/articleshow/56041967.cms

[82] A particular role in limiting the impact of the shock was played by individuals acting as "cash recyclers", essentially being paid to convert large amount of old notes into new ones. As our own evidence on consumption (Appendix D) suggests, as well as the evidence in Chodorow-Reich et al. (2019), their role does not appear to have been sufficient in shielding the Indian economy from the adverse effects of the shock.

payment options (e.g. electronic money). Indeed, we now know that almost all the old notes (99.3%) were indeed re-deposited by the deadline.

The duration of the shock Despite its magnitude, the cash crunch was a temporary phenomenon. This is to say that the period during which cash availability was a substantial constraint to conducting cash transactions was relatively short-lived. Several pieces of evidence suggest that the cash availability significantly improved in January and essentially normalized in February.

First, official statements from the post-Demonetization period indicate that the scarcity of cash was not an issue anymore by the end of January.[83] This is consistent with the aggregate behavior of cash in circulation, which grew significantly again in January 2017, suggesting that the public was able to withdraw cash from banks (Appendix Figure H.1). Furthermore, this is also consistent with the behavior of the government, which lifted most of the remaining official limitations on cash withdrawals by January 30th, 2017. In particular, it removed any ATM withdrawal limit from bank accounts. Limits had been progressively relaxed after the initial announcement, as banks started receiving the new bills. After January, the only limitation left was on withdrawal from savings accounts. Even these withdrawal limits were relatively high — Rs.50,000 per week in February 2017 —, and not necessarily binding, as households could move money to deposit accounts for withdrawals. By mid-March 2017, all limits on withdrawals from any accounts had been removed.

Second, the view that constraints on cash availability were short-lived is also consistent with the aggregate data on the use of cash in India. To start, it is important to point out that the level of cash in circulation is not necessarily informative about constraint in the use of cash for *transactions*. The reason is that only a small fraction of cash in circulation is used for transactions. This idea is extensively discussed in Rogoff (2017). While an exact quantification is difficult, Rogoff (2017) argues that less than 10% of cash in circulation in the US is actually held for transactions (Chapter 4). Furthermore, he finds evidence consistent with this qualitative pattern for several developing and developed countries, pointing out that this feature is not unique of the US. Other work provides evidence that is consistent with this point: for instance, Engert et al. (2019) studies how aggregate cash in circulation in Canada and Sweden changes over time and finds that the aggregate demand for cash does not significantly appear to be explained by the need for cash in transactions.

Withing this context, ATM withdrawals may provide a better and more direct measure of the ability to obtain new currency as needed. As a result, we examine trends in ATM withdrawal using debit card, plotted monthly in Appendix Figure H.2. We find that the total amount of cash withdrawn at ATM by debit cards returned to pre-shock (i.e. October 2016) levels around February-March 2017. This evidence supports the idea that cash was readily available for withdrawal starting in February 2017. Notice here that it is difficult to define ex-ante a clear benchmark of what we should expect in terms of timing. On the one hand, the fact that the shock increases the use of electronic payments (as we show in the paper) implies that withdrawals should not go back to the pre-shock level. On the other hand, if a lot of the cash held is actually not used for transactions, but as a store of value (Rogoff, 2017), then households may ramp up withdrawals more

[83] Several examples can be provided from the news. For instance, on January 20th it was reported that "Reserve Bank of India Governor Urjit Patel on Friday told the Public Accounts Committee (PAC) of Parliament that cash flow in the country will normalise soon." and "According to sources, Patel, who was answering queries on demonetisation and its impact, told the committee that the situation in urban areas was "almost normal" (The Sunday Guardian, January 20th 2017). Similarly, another article reports on January 25th some comments from Andhra Pradesh chief minister Chandrababu Naidu: "He said the common man's demonetisation pains were over in 60 days, adding that he was monitoring the situation daily and it was now normal in his state as well as across the country (The Information Company, January 25th 2017)."

than what would be necessary purely for transactions. With these caveats in mind, we consider our evidence as strongly supportive of the temporary nature of the shock. We come back to the use of cash by Indian households in the next sub-section.

Third, the short-lived nature of constraints on cash availability is also visible from data on Internet searches. We use searches as a way to elicit public perceptions of cash availability, as a complement to the more direct measures (but potentially difficult to interpret) measures of cash availability just discussed. Appendix Figure H.3 reports monthly data (from 09/2016 to 07/2017) of Google searches for several key words that could be associated with the shock. For instance, we collect data on searches on the words "Cash" or "ATM line," among others. Data is obtained by Google Trends, and the index is normalized by Google to be from 0 to 100, with the value of 100 assigned to the day with the maximum number of searches made on that topic. Across all the panels, we find that Google searches that are related to the Demonetization spiked in November, remained high in December, but then significantly dropped in January, before returning to preshock levels in February. One exception is the search on "ATM Cash withdrawal limit today" which reached its maximum on January 31, 2017. This is consistent with the fact that January 31, 2017 was the date when most limits on ATM withdrawals were lifted by the RBI. Altogether, this information also points to relatively short-lived constraints on cash availability, and lines up with the timing discussed in the main paper.

Therefore, while large in magnitude, the shock to cash induced by the Demonetization was relatively short in terms of duration. In general, cash scarcity was very high in November and December 2016, the general conditions improved significantly over January 2017, and the situation had normalized starting with February 2017.

Other aggregate effects While the shock generated a temporary shortage in cash, one concern for our analysis is that it may have also affected other aspects of the Indian economy in a persistent way, driving long-term adoption independently from our key mechanism. To start, it is important to point out that this type of issue is exactly what led us to implement the key empirical tests of our model using disaggregated data. In fact, the presence of other aggregate changes in the Indian economy affecting firms' propensity to adopt does not necessarily affect the analyses using district-level variation, where aggregate changes are net out by the presence of time effects. In other words, our reduced-form tests naturally relax the type of identification assumption necessary to test our model, and are robust to a broader class of potential confounding factors. To be clear, this statement does not imply that no aggregate shock can affect our reduced-form analysis. We refer to Section 4.4 for a careful discussion of identification in our reduced-form analysis (and the various tests presented to rule out alternative interpretations).[84]

Despite the advantages of our setting, we also want to directly examine the leading concerns. First, we consider the role of economic uncertainty during the Demonetization period (Bloom, 2009). In fact, while the Demonetization led to a temporary shortage of cash, it may have significantly and persistently affected the level of uncertainty about the condition in the economy (or the supply of cash), which in turn may have affected the relative incentive to adopt electronic payments.[85]

Our analysis highlights how the increase in uncertainty is mostly temporary, and therefore it is not likely

[84] Furthermore, we also recognize that the presence of other aggregate shifts caused by the Demonetization may affect some of our counterfactual analyses. We discuss this issue and examine how the nature of the shock affects our estimates from the structural model in Section 5.

[85] A separate dimension of uncertainty is the uncertainty about the nature of the technology. We see this aspect as closely related to the learning mechanism, since the presence of uncertainty around the new technology represents a natural requirement to justify the importance of learning from others. See Appendix F for more discussion on this issue.

to affect long-term adoption in a way that would affect our results. To start, we examine aggregate measures of uncertainty in India. In the two panels of Appendix Figure H.11, we report the plot of two measures of uncertainty constructed by researchers at the Reserve Bank of India (Priyaranjan and Pratap, 2020).[86] These figures measure uncertainty using a text-based algorithm that tries to extract information from newspapers (panel A) or Google Search data (panel B).[87] Across both series, we find an increase in uncertainty right around November 2016, with the peak experienced in either January or February 2017 depending on the series.[88] To interpret these findings, we also need to highlight that in the second part of January there were a lot of actions by the government to lift cash limitations, and these interventions will mechanically increase the measured level of uncertainty. However, after this period, the level of uncertainty in both series go back to the same (noisy) pattern that characterized the series in the pre-Demonetization period.[89] Therefore, the Demonetization may have increased uncertainty, but only temporarily.

One concern with this analysis is that it does not directly speak to the uncertainty about the shock to cash. For instance, Indian households may have been worried of further reduction in cash after January, which may have affected their adoption incentives. Two pieces of evidence reject this specific concern. First, it is worth pointing out that the Google Trends analysis already mentioned earlier in this Section also appears inconsistent with the presence of concerns of future policies.

Second, we also find no evidence of widespread beliefs about the possibility of a new policy. To examine this question specifically, we started by hiring a research assistant to search through Indian newspaper articles available on ProQuest TDM. We search over the articles published in the three months following the Demonetization (i.e. Nov 8, 2016 to Feb 8, 2017). We constrained our search on articles that mentioned "Demonetization" and its synonyms.[90] This yielded a total of approximately 31,000 articles. We then developed a list of bigrams that could be associated with future occurrences of an event [91] and scrapped the sentences in articles and their headlines in which there is a reference of Demonetization (or its associated synonyms) with +/- 10 words of these bi-grams. We then manually examined this set of about 1,200 pieces of text.[92] This further screen left us with only 16 articles mentioning the possibility of a future Demonetization, which we then read in full. At the end, starting from the sample of approximately 31,000 articles discussing the Demonetization, only 7 explicitly mentioned the possibility of future policies similar to the Demonetization.[93] This evidence supports the view that the likelihood of another set of restrictions

[86]We kindly thank the authors to share the data from the paper.

[87]Priyaranjan and Pratap (2020) also reports a third series based on Lexicon data. Rather than measure uncertainty, this series seems to focus more on corporate sentiment. Consistent with this interpretation, we do not see any detectable change in this series around Demonetization. As a result, our conclusions would not change if we also had considered this series. If interested, time-series of this measure is available in Priyaranjan and Pratap (2020).

[88]One general caveat with this analysis is that measures of uncertainty for India (similar to other developing countries) are generally much noisier than the same proxies for US.

[89]We have also further investigated the higher reported uncertainty in February 2017 in the Google search series. Examining the underlying data, we were told by the authors that the increase in February is mostly driven by search related to fiscal policy. In particular, these searches are related to the government's decision to present its annual budget in February 2017, shedding the old practice of presenting the budget in March.

[90]Specifically, we searched "Demoneti?ation" to allow for different ways Demonetization is spelled. We also allowed for various synonyms for the Demonetization that were used by the press including *the event, the policy, cash ban, the move, the announcement, the initiative, the ban, demonetize, demonetizing, demonetise, demonetising*.

[91]Specifically, we look for: might again, could happen, happen again, do again, it again, never happen, never again, government may, RBI may, likely to, unlikely to, likely never, likely will, possible that, once again, another round, new currency, new notes, 2,000 new, 2000 new, thousand new.

[92]Specifically, we had one research assistant and one of us separately read through all the 1,200 parts of the articles (i.e. the set of words around the found keyword, as described above) and classified the articles that we believed fit the criteria of mentioning future policies.

[93]Of the remaining 9 articles excluded when reading them in full: 2 took a negative stand on the future possibility of such policies, 4 were opinion-pieces suggesting what the government should do another round of Demonetization, one mentioned a government statement stating it lack of intention to repeat the Demonetization, and 2 articles framed the statement as a

on cash was perceived as low by the media.

In general, concerns about cash goes back to pre-shock level around the end of January 2017, therefore rejecting that the shock led to a persistent increase in uncertainty. Altogether, while as expected the Demonetization led to an increase in uncertainty, the temporary nature of this economic force suggests that this mechanism is unlikely to play a leading role in explaining our results.

A second concern about our setting is that Demonetization — on top of increasing the size of the electronic payment network — could have also affected the long-run incentives to use mobile wallets by reducing the value of cash as a store-of-value. The idea is that the policy may have changed the preferences of the Indian population toward cash holdings, for reason unrelated to the transactional value of cash versus electronic payments.[94] However, data on the use of cash appears at odds with this hypothesis. First, the evidence on ATM cash withdrawal discussed earlier (Figure H.2) already appears inconsistent with this hypothesis. With all the caveats discussed earlier in mind, aggregate ATM withdrawals in India came back to their October levels shortly after January 2017, suggesting that households and firms did not persistently move away from cash.

Second, we find that the fraction of aggregate holdings of liquid wealth in the form of cash also recovered back to the pre-shock average. As we discussed above, aggregate cash in circulation is more likely to capture aggregate demand of cash for store-of-value rather than its transactional value (Engert et al., 2019). To examine this question, we plot the amount of cash in circulation relative to the total money supply (M3) obtained from the Reserve Bank of India (Appendix Figure H.12).[95] Not surprisingly, the share of cash in circulation declined significantly in November and December 2016 because of the Demonetization. However, over the long run, the ratio went back to its long-term average relatively quickly: by early 2018, cash in circulation was back about 13% of liquid wealth, and stayed at that level after that. As discussed before, it is hard to state a clear benchmark about the expected timing of this response under the assumption that the preference for cash did not change. In general, we think of this reversal to be actually very fast, since cash in circulation is a stock variable and therefore it would require some time for households to go back to the pre-shock level immediately. Altogether, similar to Lahiri (2020), we conclude that the Demonetization did not affect the preferences for holding cash.

A.2 Subsequent policy interventions

While the initial objective of the government was not to foster a shift towards electronic payments, the increase in electronic payments following the Demonetization did not go unnoticed. As a result, the government decided to intervene more actively in this space.

On top of generic announcements from top politicians, the government and the RBI put into place some interventions in the area of electric payments following the Demonetization. First, the government actively supported the adoption of traditional electronic payment technologies by trying to lower the adoption costs of point-of-sales (POS) system, in particular for small businesses. One example of this type of program was the grants that were provided by the National Bank for Agriculture and Rural Development (NABARD) to support the acquisition of POS machines in small villages.[96] Second, the government partnered with

question rather than an expectation.

[94] For instance, wealthy individuals may have now realized the risk of holding high cash balances or the implicit costs of this strategy.

[95] We use M3 at the denominator to provide a high-frequency measure of aggregate money in the economy. However, results are unchanged if we scale cash in circulation by GDP.

[96] https://indianexpress.com/article/business/business-others/demonetisation-nabard-cashless-economy-debit-cards-4417938/

several other organizations to provide discounts on their products when payments were made electronically. The main discounts involved gas and railroads. For instance, the government partnered with Indian Oil Corp, Bharat Petroleum, and Hindustan Petroleum to give a 0.75% discount to consumers if they paid electronically. For railroads, the incentives ranged from a small discount on ticket acquired with electronic payments —- generally up to 0.5% — to free accident insurance for travelers.[97] These policies were announced on December 8, 2016 and implemented either immediately or by the end of the month (for instance, the gas station incentives were implemented on December 9, and railroad ticket incentives were implemented on January 1st 2017).

Similar to our earlier discussion, the presence of the government's response should not automatically be a problem for our reduced-form analysis. In fact, this approach differences out any aggregate change in policy during our period. Several specific features of these policies reinforce the idea that our models should be useful to control for these factors. First, from a detailed analysis of the subsequent policy changes, we found no evidence that any of these interventions were designed formally or informally to target specific areas more affected by the cash contraction. Second, as we discuss more in detail below, we do not find any systematic differences in the response to policy announcements between more or less affected areas. This aspect (combined with the balance of our treatment on observable) is important because — even if the policies did not target specific areas — they could still have heterogeneous effects.

Therefore, our approach is in principle well suited to examine the impact of the cash contraction, conditional on aggregate changes. To further provide evidence consistent with this hypothesis, we want to also highlight two other important features of these policies. First, most of these interventions targeted more traditional electronic payment technologies, and not fintech platforms, and therefore they are — if anything — going to bias us towards finding no effect on our mobile wallet technology. One important example is the policy put forward to foster POS terminal adoption. In fact, POS terminals are the basic infrastructure for conducting credit cards and debit cards' transactions but are completely irrelevant for using the technology covered by our data. In this context, the provision of subsidies to acquire POS terminals — if successful — would just reduce the expected response in terms of mobile wallet, therefore biasing this response towards zero. This point goes beyond POS terminals: if the government's push did not directly target our technology, but did target related alternatives, it should attenuate our estimates.

Second, it is unclear whether these policies were effective at all. To be clear, a full policy evaluation of these interventions is outside the scope of the paper. However, we can use our data to examine whether these specific policies affected the adoption of our mobile wallet technology and whether this response was somehow correlated with the initial exposure to the shock, which is the specific concern with our analysis.

To start, we can directly use high-frequency aggregate data to check whether there was any structural change in the use of electronic payments around the time a policy was announced (or introduced). Figure 1 in the paper can already shed some evidence on the effectiveness of these policies by themselves in driving e-wallet adoption. The figure shows that the growth rate in the amount transacted on our e-wallet platform did not change significantly to these policy announcements, suggesting that these policies were not a major driving force behind the growth in adoption. While positive for most of this period, the growth rate has been declining at a roughly constant rate around the announcement or implementations, suggesting that in aggregate these policies did not significantly affect the adoption of e-wallet.

Next, we move past the aggregate evidence, and exploit variation across districts. This is relevant because

[97] https://economictimes.indiatimes.com/industry/transportation/railways/soon-you-will-be-rewarded-for-cashless-booking-of-railway-tickets/articleshow/61954004.cms?from=mdr.

the aggregate evidence could mask considerable heterogeneous responses across districts. For this reason, we conduct a cross-sectional test by analyzing the growth rate in payment activity across districts with different levels of exposure in a narrow-window (i.e two-weeks) around these policy changes. The results are reported in Appendix Table H.9.[98]

To provide a benchmark, we start by presenting the relationship between our treatment and the growth rate around our main policy event (i.e. the Demonetization). Just to be clear, this is essentially a replication of our main result employing this alternative specification and with higher frequency data. We then contrast to what happens after the announcement of the new government policies (column 2) and their implementation (column 3). Unlike around our main event (where we replicate our main result), we do not see a change in growth rates around the government intervention that is correlated with our treatment, suggesting that these policies did not have any incremental effect in the use of electronic payments in ex-ante more exposed areas.[99] Therefore, even to the extent that the policy had heterogeneous effects across areas, this heterogeneity does not appear to be correlated with our treatment.

While this discussion should not be interpreted as a full policy evaluation of the government's response, it does more narrowly suggest that concerns about the role of policy intervention as a confounding factor in our analysis are quite limited: in general, we do not see around the policy announcement or introduction any significant change in adoption, either in aggregate or across districts.

A.3 The electronic wallet technology

The main focus of the paper is on the adoption of a specific electronic wallet. This section provides further details on this technology.

Our data provider was one of the main fintech company active in India at the time, and the largest player in the provision of electronic wallet payment services during Demonetization. In terms of data, the company specifically shared with us information on their payment product targeting small and medium-sized retailers. Specifically, the company has shared information at retailer-level — number of transactions and amount per week — covering the quasi-totality of their activity within this product-type. This information allows us to identify the time when the company joins the platform, as well as its subsequent use. On top of information on the use of the platform, the data also contains the business's location and type of merchant. One implication of our sample is that issues related to retailer market power are probably not first order. In fact, while there are firms in our sample that are likely competitors with each other (i.e. firms within the same location and merchant type), the typical firm in our sample is relatively small in size.

In terms of technology, the company allows individuals and businesses to undertake transactions with each other using only their mobile phone. To use the service, a customer would normally need to download an application and link their bank account to the application. However, in 2016 the company also established a new service that allows customers to make payments without the need of Internet or a smart phone. Thus, the technology offers multiple ways to complete a transaction. First, customers can scan the merchants' unique QR code in the application installed on their smart phones. Second, instead of scanning the QR code, customers can enter the mobile number of the merchant. In this case, the merchant would receive a unique code from the company, which is then used by the customer to complete the transaction. Third, if a

[98] Specifically, we estimate $\Delta y_d = \beta \text{Exposure}_d + \gamma X_d + \epsilon_d$, where Δy_d is the two-week symmetric growth rate in transactions conducted on the platform around the policy date (i.e. the growth between the week before and two after), and X_d are the set of covariates from our baseline specification (15) in the paper. The policy date is reported in the header of the table.

[99] Furthermore, the estimates in Columns 2 to 3 should be considered an upper bound for the true effects from these two new government policies since they may in part capture the "long-wave" of the Demonetization on adoption growth.

smart phone or mobile Internet are not available, customers can call a toll-free number and ask the wallet company to complete the transaction using the cell-phone number of the merchant. To use this feature, customers needed to be enrolled through a one-time verification process.

In general, the requirements to be able to transact — i.e. to have a phone and a bank account — were not particularly binding for India in 2016. In terms of bank accounts, India has an extremely high penetration of formal bank accounts, which is high also relative to some developed countries. This situation is in part the results of some policy interventions put in place in the previous decade. The most prominent example is the Pradhan Mantri Jan Dhan Yojna, which was launched in 2014 and led to more than 200 million new bank account openings (Agarwal et al., 2017). Our representative household survey (CMIE) confirms this idea: on average, we find that 96% of households have a bank account in a district. Altogether, the access to bank accounts was not a significant constraint for households wanting to use our technology.[100]

Phone penetration was also relatively high, similar to other developing countries. The CMIE's household survey confirms this finding: in our data, 93% have a mobile phone. While we do not know how many of these phones were smart phones, two points need to be made. First, as discussed above, households also had the option to use our company's payment system without Internet. Second, the 2G penetration was very high in India during this period, with a coverage of 93.5% (World Bank). The bottom line is that phone access — similar to bank accounts — was also not a particularly constraint for consumers willing to access mobile wallets.

Once a payment has been received, retailers can transfer the money from their electronic wallet to a traditional bank account. Therefore, the technology is in many respect similar to a credit card or other more traditional electronic payment systems. However, relative to these other electronic payment technologies, adoption costs are much lower, since merchants and consumers can access the electronic wallet almost instantaneously, without the need for anything more than a phone and a bank account. In particular, from the standpoint of the retailers, the mobile wallet does not require the acquisition of a POS.

On top of fixed costs, variable costs of this technology are also very limited, in particular for small merchants. Merchants using the digital wallet are classified by the provider into three segments: small, medium and large. Small merchants have lower limits on the amount they can transact but pay no transaction costs. Medium-size merchants can transfer money to their bank account at midnight every day up to a certain limit. Large merchants can transact any amount but pay a percentage of the transfer amount as a fees. Our data only covers small and medium-size merchants. From discussions with the company, large merchants tend to have more personalized contracts that can bundle different services and payment options together.

Finally, we discuss briefly the competitive landscape for wallet payment systems in India. As mentioned before, our company was by far the largest provider of this service during the period we study, and could be considered the de-facto monopolist for most of our sample. However, this does not necessarily imply that competition between platforms is not important. For instance, even if competition was not a first order concern at an early point, the future threat of competition may still play a role on the way the platform is structured and the response of consumers. However, there are two key aspects of competition that are worth discussing. First, the presence of actual or perceived future competition is likely to — if anything — reduce our estimate of the response to the shock.[101] Therefore, we do not think that the presence of potential competitors can spuriously generate any of our findings and threaten internal validity. Second, situations of near-monopoly are common when studying the early life of a new technology. When considering the

[100] Additionally, our company provides options to deposit money into the mobile wallet without the need of a bank account.

[101] To the extent that there are other providers that can also offer a similar product (now or in the near future), this implies that there are alternatives to our platform and should make — all else equal — our response to the shock smaller.

introduction of a new product or service, the first mover company is likely to be the de-facto monopolist for some time. Therefore, while we recognize that these contextual factors need to be carefully considered when trying to extrapolate our results outside our specific setting, we also think that the competitive landscape in the empirical setting we consider shares some of typical traits of technologies at early stages of adoption.

B Theory

This appendix provides proofs for the results reported in Section 3, as well as for the fixed cost model discussed in Section 4.4.

B.1 Proofs of Lemma 1 and Results 1 and 2

We start with the following preliminary Lemma, which is used in the proof of Lemma 1 from the main text.

Lemma 2. *The conditional distribution of M_t is given by:*

$$\forall s \geq t, \quad M_s | M_t \sim N(\mu_{s|t}, \sigma^2_{s|t})$$

$$\mu_{s|t} = \begin{cases} e^{-\theta(s-t)} M_t + (1 - e^{-\theta(s-t)}) M^c & \text{if} \quad t \leq T \text{ and } t \leq s \leq T \\ e^{-\theta(T-t)} M_t + (1 - e^{-\theta(T-t)}) M^c & \text{if} \quad t \leq T \text{ and } T \leq s \\ M_t & \text{if} \quad t \geq T \end{cases},$$

$$\sigma^2_{s|t} = \begin{cases} (1 - e^{-2\theta(s-t)}) \dfrac{\sigma^2}{2\theta} & \text{if} \quad t \leq T \text{ and } t \leq s \leq T \\ (1 - e^{-2\theta(T-t)}) \dfrac{\sigma^2}{2\theta} + (s-T)\sigma^2 & \text{if} \quad t \leq T \text{ and } T \leq s \\ (s-t)\sigma^2 & \text{if} \quad t \geq T \end{cases}.$$

Proof of Lemma 2. Note that for $t \leq T \leq s$, we can write:

$$M_s = M_t + (M_T - M_t) + (M_s - M_T).$$

The increment $D_T = (M_T - M_t)$ is independent from the increment $(M_s - M_T)$. Moreover, for any $t \leq u \leq T$,

$$dD_u = \theta(M^c - M_t - D_u)du + \sigma dZ_u,$$

implying that $\mathbb{E}_t[D_T] = (1 - e^{-\theta(T-t)})(M^c - M_t)$ and $\mathbb{V}_t[D_T] = (1 - e^{-2\theta(T-t)}) \dfrac{\sigma^2}{2\theta}$. □

We now prove Lemma 1 from the main text. We omit the firm indices to simplify notation.

Proof of Lemma 1. The Bellman equation for the value of a firm with technology choice $x_t = e$ is:

$$\begin{aligned} V_t(e, M_t, X_t) \quad = \quad & \max_{\tilde{k} \in [0,k]} \Big\{ \Pi(e, M_t, X_t) dt \\ & + \quad \tilde{k} dt (1 - r dt) \mathbb{E}_t \left[V_t(c, M_{t+dt}, X_{t+dt}) \right] \\ & + \quad (1 - \tilde{k} dt)(1 - r dt) \mathbb{E}_t \left[V_t(e, M_{t+dt}, X_{t+dt}) \right] \Big\} \end{aligned}$$

Substituting $dV_t(e, M_t, X_t) = V_t(e, M_{t+dt}, X_{t+dt}) - V_t(e, M_t, X_t)$, re-organizing the equation above, and omitting terms of order $(dt)^2$ and higher, we obtain:

$$rV_t(e, M_t, X_t)dt = \Pi(e, M_t, X_t)dt + \mathbb{E}_t\left[dV_t(e, M_t, X_t)\right] - \max_{\tilde{k}\in[0,k]} \tilde{k}B_t(M_t, X_t)dt,$$

where:

$$B_t(M_t, X_t) \equiv V_t(e, M_t, X_t) - V_t(c, M_t, X_t).$$

Likewise,

$$rV_t(c, M_t, X_t)dt = \Pi(c, M_t, X_t)dt + \mathbb{E}_t\left[dV_t(c, M_t, X_t)\right] + \max_{\tilde{k}\in[0,k]} \tilde{k}B_t(M_t, X_t)dt,$$

Thus, in general:

$$\begin{aligned}\tilde{k}_t(e, M_t, X_t) &= \arg\max_{\tilde{k}\in[0,k]} -kB_t(M_t, X_t),\\ \tilde{k}_t(c, M_t, X_t) &= \arg\max_{\tilde{k}\in[0,k]} kB_t(M_t, X_t).\end{aligned}$$

To facilitate exposition, we assume that if $B_t(M_t, X_t) = 0$ then $\tilde{k}_t(e, M_t, X_t) = 0$ and $\tilde{k}_t(c, M_t, X_t) = k$. This is without loss of generality, since Frankel and Burdzy (2005) (footnote 9, p.13) show that, even when the optimal arrival rate is a correspondence, $B_t(M_t, X_t)$ is equal to 0 only on a measure-zero set of states.

In order to derive the expression for B_t, we first combine the two Bellman equations, with the fact that $\tilde{k}_t(e, M_t, X_t) + \tilde{k}_t(c, M_t, X_t) = k$ to obtain:

$$\mathbb{E}_t\left[dV_t(e, M_t, X_t) - dV_t(c, M_t, X_t)\right] = (r+k)B_t(M_t, X_t)dt - \Delta\Pi(M_t, X_t)dt \tag{20}$$

Multiplying the left-hand side by $e^{-(r+k)s}$ and integrating by parts from $s = 0$ to $s = S$, we obtain:

$$\begin{aligned}&\mathbb{E}_t\left[\int_{s=0}^{S} e^{-(r+k)s}\left(dV_{t+s}(e, M_{t+s}, X_{t+s}) - dV_{t+s}(c, M_{t+s}, X_{t+s})\right)ds\right]\\ =\;& (r+k)\mathbb{E}_t\left[e^{-(r+k)S}B_{t+S}(M_{t+S}, X_{t+S})\right] - B_t(M_t, X_t)\\ +\;& (r+k)\mathbb{E}_t\left[\int_{s=0}^{S} e^{-(r+k)s}B_{t+s}(M_{t+s}, X_{t+s})ds\right]\end{aligned} \tag{21}$$

Multiplying the right-hand side of equation 20 by $e^{-(r+k)s}$, integrating from $s = 0$ to $s = S$, and comparing to Equation (21), we obtain:

$$\mathbb{E}_t\left[\int_{s=0}^{S} e^{-(r+k)s}\Delta\Pi(M_{t+s}, X_{t+s})ds\right] = B_t(M_t, X_t) - \mathbb{E}_t\left[e^{-(r+k)S}B_{t+S}(M_{t+S}, X_{t+S})\right] \tag{22}$$

In order to conclude we need to establish that for all $t \geq 0$,

$$\lim_{S\to+\infty} \mathbb{E}_t\left[e^{-(r+k)S}B_{t+S}(M_{t+S}, X_{t+S})\right] = 0.$$

We have the two following bounds on the value of the firm:

$$
\begin{aligned}
V_t(x_t, M_t, X_t) &\leq \mathbb{E}_t\left[\int_{s=0}^{+\infty} e^{-rs} \max\left(\Pi(e, M_{t+s}, X_{t+s}), \Pi(c, M_{t+s}, X_{t+s})\right) ds\right]. \\
V_t(x_t, M_t, X_t) &\geq \mathbb{E}_t\left[\int_{s=0}^{+\infty} e^{-rs} \min\left(\Pi(e, M_{t+s}, X_{t+s}), \Pi(c, M_{t+s}, X_{t+s})\right) ds\right].
\end{aligned}
$$

and therefore:

$$
|B_t(M_t, X_t)| \leq \mathbb{E}_t\left[\int_{s=0}^{+\infty} e^{-rs}|\Delta\Pi(M_{t+s}, X_{t+s})|ds\right].
$$

Fix $v > T$. For all $s \geq 0$, we have:

$$
|\Delta\Pi(M_{v+s}, X_{v+s})| \leq |M_v| + |\tilde{M}_{v+s}| + C + M^e. \tag{23}
$$

where:

$$
\forall s \geq 0, \quad \tilde{M}_{v+s} \equiv M_{v+s} - M_v.
$$

For any $0 \leq s$, $\tilde{M}_{v+s}$ is conditionally normal with mean 0 and variance $s\sigma^2$. So $|\tilde{M}_{v+s}|$ follows the corresponding half-normal distribution, and therefore:

$$
\mathbb{E}_v\left[|\tilde{M}_{v+s}|\right] = s\sigma^2\sqrt{\frac{2}{\pi}}
$$

Therefore,

$$
\mathbb{E}_v\left[\int_{s=0}^{+\infty} e^{-rs}|\Delta\Pi(M_{v+s}, X_{v+s})|ds\right] \leq \frac{C + M^e + |M_v|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}}
$$

so that:

$$
|B_v(M_v, X_v)| \leq \frac{C + M^e + |M_v|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}}
$$

Fix $t > 0$ and, without loss of generality, let $S > T - t$. Then $t + S > T$, so:

$$
\begin{aligned}
|B_{t+S}(M_{t+S}, X_{t+S})| &\leq \frac{C + M^e + |M_{t+S}|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}} \\
&\leq \frac{C + M^e + |M_t| + |\tilde{M}_{t+S}|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}}
\end{aligned}
$$

where again, $\tilde{M}_{t+S} \equiv M_{t+S} - M_t$.

If $t > T$, then $\tilde{M}_{t+S}$ is conditionally normal with mean 0 and variance $S\sigma^2$. Therefore,

$$
\mathbb{E}_t\left[|B_{t+S}(M_{t+S}, X_{t+S})|\right] \leq \frac{C + M^e + |M_t|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}} + \frac{\sigma^2}{r}\sqrt{\frac{2}{\pi}}S,
$$

which implies that

$$
\lim_{S\to+\infty} \mathbb{E}_t\left[e^{-(r+k)S}|B_{t+S}(M_{t+S}, X_{t+S})|\right] = 0.
$$

If $t \leq T$, then $\tilde{M}_{t+S}$ is conditionally normal with mean and variance:

$$\begin{aligned}
\mu_{t,t+S} &= (1-e^{-\theta(T-t)})(M^c - M_t) \\
\sigma^2_{t,t+S} &= (t+S-T)\sigma^2 + \frac{\sigma^2}{2\theta}(1-e^{-2\theta(T-t)})
\end{aligned}$$

Then $|\tilde{M}_{t+s}|$ is conditionally half-normal, so that:

$$\begin{aligned}
\mathbb{E}_t\left[|\tilde{M}_{t+S}|\right] &= \sigma_{t,t+S}\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2_{t,t+S}}{2\sigma^2_{t,t+S}}} + \mu_{t,t+S}\left(1-2F\left(-\frac{\mu_{t,t+S}}{\sigma_{t,t+S}}\right)\right) \\
&\leq \sigma_{t,t+S}\sqrt{\frac{2}{\pi}} + |\mu_{t,t+S}|
\end{aligned}$$

where F is the standard normal CDF. Thus,

$$\mathbb{E}_t\left[|B_{t+S}(M_{t+S}, X_{t+S})|\right] \leq \frac{C + M^e + |M_t|}{r} + \frac{\sigma^2}{r^2}\sqrt{\frac{2}{\pi}} + \frac{1}{r}\left(\sigma_{t,t+S}\sqrt{\frac{2}{\pi}} + |\mu_{t,t+S}|\right).$$

Since $\lim_{S\to+\infty} \sqrt{t+S-T}e^{-(r+k)S} = 0$, we have:

$$\lim_{S\to+\infty} \frac{e^{-(r+k)S}}{r}\left(\sigma_{t,t+S}\sqrt{\frac{2}{\pi}} + |\mu_{t,t+S}|\right) = 0,$$

and therefore $\lim_{S\to+\infty} \mathbb{E}_t\left[e^{-(r+k)S}|B_{t+S}(M_{t+S}, X_{t+S})|\right] = 0$, concluding the proof. □

In order to obtain Results 1 and 2, we start by establishing that in any equilibrium the adoption rules (and therefore the value functions) must be symmetric across firms. Indeed, consider a candidate set of equilibrium strategies $\left\{a^c_{i,t}\right\}_{t\geq 0, i\in[0,1]}$, and define:

$$a^c_t(M_t, X_t) \equiv \int_i a^c_{i,t}(M_t, X_t)di. \tag{24}$$

Then we can rewrite the law of motion for X_t induced by these strategies as:

$$dX_t = (a^c_t(M_t, X_t) - X_t)kdt. \tag{25}$$

In other words, $a^c_t(M_t, X_t)$ is a sufficient statistic for the effect of other firm's actions on the aggregate law of motion of X_t. As a result, when computing the best response to $\left\{a^c_{i,t}\right\}_{t\geq 0, i\in[0,1]}$ using Equation (5) and (7), all firms will use the same law of motion for X_t. Their best responses will be therefore be identical, so that the equilibrium must be symmetric.

We then establish the following additional Lemma. It states that there are are two strict dominance regions in the model: $M_t \geq \overline{\Phi}_t(X_t)$, where cash strictly dominates; and $M_t < \underline{\Phi}_t(X_t)$, where electronic payments strictly dominate. The boundaries $\overline{\Phi}_t(.)$ and $\underline{\Phi}_t(.)$ are parallel; they coincide, if and only if, $C = 0$.

Lemma 3 (Strict dominance regions)**.** *For any sequence of adoption rules a, we have:*

$$B_t(M_t, X_t; a_{|t}) = A_{M,t}\left(\underline{M}_t - M_t\right) + A_X X_t + A_N N_t(M_t, X_t; a_{|t}),$$

$$N_t(M_t, X_t; a_{|t}) = \mathbb{E}_t\left[\int_t^{+\infty} e^{-(r+k)(s-t)} a_s(M_s, X_s) ds | a_{|t}\right],$$

$$A_{M,t} \equiv \begin{cases} \dfrac{1}{r+k+\theta} + \dfrac{\theta e^{-(r+k+\theta)(T-t)}}{(r+k+\theta)(r+k)} & \textit{if} \quad t \leq T \\ \dfrac{1}{r+k} & \textit{if} \quad t > T \end{cases},$$

$$\underline{M}_t \equiv \begin{cases} M^c - A_{M,t}^{-1}\dfrac{M^c - M^e}{r+k} & \textit{if} \quad t \leq T \\ M^e & \textit{if} \quad t > T \end{cases},$$

$$A_X \equiv \frac{1}{r+2k}C,$$

$$A_N \equiv kA_X.$$

The value of adoption given a strategy profile is bounded as follows:

$$A_{M,t}(\underline{\Phi}_t(X_t) - M_t) \leq B_t(M_t, X_t; a_{|t}) \leq A_{M,t}(\overline{\Phi}_t(X_t) - M_t), \tag{26}$$

where:

$$\begin{aligned} \underline{\Phi}_t(X) &\equiv \underline{M}_t + \frac{A_X}{A_{M,t}}X, \\ \overline{\Phi}_t(X) &\equiv \underline{\Phi}_t(X) + \frac{A_X}{A_{M,t}}\frac{k}{r+k}. \end{aligned} \tag{27}$$

Any equilibrium sequence of adoption rules must satisfy:

$$a_t(M_t, X_t) = \begin{cases} 1 & \textit{if} \quad M_t \leq \underline{\Phi}_t(X_t), \\ 0 & \textit{if} \quad M_t > \overline{\Phi}_t(X_t). \end{cases} \tag{28}$$

Proof of Lemma 3. Using the result of Lemma 1, we can write the value of adoption as:

$$\begin{aligned} B_t(M_t, X_t; a_{|t}) &= B_t^{(1)}(M_t) + C \times B_t^{(2)}(M_t, X_t; a_{|t}), \\ B_t^{(1)}(M_t) &\equiv \mathbb{E}_t\Big[\int_t^{+\infty} e^{-(r+k)(s-t)}\left(M^e - M_s\right) ds\Big], \\ B_t^{(2)}(M_t, X_t; a_{|t}) &\equiv \mathbb{E}_t\Big[\int_t^{+\infty} e^{-(r+k)(s-t)} X_s ds \mid a_{|t}\Big]. \end{aligned} \tag{29}$$

Computation shows that:

$$B_t^{(1)}(M_t) = \begin{cases} -\dfrac{M^c - M^e}{r+k} + \left(\dfrac{1}{r+k+\theta} + \dfrac{\theta e^{-(r+k+\theta)(T-t)}}{(r+k+\theta)(r+k)}\right)(M^c - M_t) & \text{if} \quad t \leq T \\ \dfrac{M^e - M_t}{r+k} & \text{if} \quad t > T. \end{cases}$$

$$= A_{M,t}\left(\underline{M}_t - M_t\right).$$

For $B^{(2)}$, first note that for all $s \geq t$,

$$X_s \quad = k \int_t^s e^{-k(s-u)} a_u(M_u, X_u) du + e^{-k(s-t)} X_t.$$

Therefore:

$$\begin{aligned}
& B_t^{(2)}(M_t, X_t; a_{|t}) \\
=& \left(\int_t^{+\infty} e^{-(r+2k)(s-t)} ds \right) X_t + k \mathbb{E}_t \Big[\int_t^{+\infty} e^{-(r+k)(s-t)} \left(\int_t^s e^{-k(s-u)} a_u(M_u, X_u) du \right) ds \mid a_{|t} \Big] \\
=& \frac{1}{r+2k} X_t + k \mathbb{E}_t \Big[\int_t^{+\infty} \int_t^s e^{-(r+2k)(s-t)} e^{k(u-t)} a_u(M_u, X_u) du ds \mid a_{|t} \Big]
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \int_{s=t}^{+\infty} \int_{u=t}^s e^{-(r+2k)(s-t)} e^{k(u-t)} a_u(M_u, X_u) du ds \\
=& \int_{u=t}^{+\infty} \int_{s=u}^{+\infty} e^{-(r+2k)(s-t)} e^{k(u-t)} a_u(M_u, X_u) du ds \\
=& \frac{1}{r+2k} \int_t^{+\infty} e^{-(r+k)(u-t)} a_u(M_u, X_u) du,
\end{aligned}$$

so that:

$$B_t^{(2)}(M_t, X_t; a_{|t}) = \frac{1}{r+2k} X_t + \frac{k}{r+2k} N_t(M_t, X_t; a_{|t}).$$

The bounds for B are then obtained by noting that, for any strategy profile a:

$$0 \leq N_t(M_t, X_t; a_{|t}) \leq \frac{1}{r+k}, \tag{30}$$

where the upper bound corresponds to the value obtained if all firms choose adoption of electronic money in all future dates and states ($a_t(M_t, X_t) = 1$ for all M_t, X_t), while the lower bound is obtained if all firms move to cash in all future dates and states ($a_t(M_t, X_t) = 0$ for all M_t, X_t).

When $M_t > \overline{\Phi}_t(X_t)$, adopting electronic money is strictly dominated because it yields negative adoption benefits B even if other firms choose to adopt it in all future dates and states. So it cannot be part of a Nash equilibrium. Likewise, when $M_t \leq \overline{\Phi}_t(X_t)$, adopting cash is strictly dominated because it yields negative adoption benefits B even if no firms adopt electronic money in all future dates and states. $\square$

We now turn to the proof of Result 1. Appendix Table H.19 shows how to define the objects of the more general framework of Frankel and Burdzy (2005) in terms of our model objects. Thus, our model is a particular case of their framework. We show that our model primitives satisfy Assumptions A0 to A6 in their paper, which are sufficient conditions for the results which we state in the main text.[102] Only Assumption A5 below requires an explicit proof; the other follow from simple computations.

[102] Assumption $A0$ is not explicitly stated by Frankel and Burdzy (2005) but is required by their results.

Assumption A0 (Lipschitz relative payoff). *The relative payoff flow in mode 1 is Lipschitz in W and X, with constants $\beta = C, \quad \overline{\alpha} = 1$; that is, for all W, W', X, X',*

$$
\begin{aligned}
D(W,X) - D(W,X') &\le \beta|X' - X|, \\
D(W,X) - D(W',X) &\le \overline{\alpha}|W' - W|.
\end{aligned}
\tag{31}
$$

Assumption A1 (Bounded switching rates). *The switching rates $k^1 = \tilde{k}(e, M_t, X_t)$ and $k^2 = \tilde{k}(c, M_t, X_t)$ must satisfy:*

$$
\begin{aligned}
&k^m \in \left[\underline{K}^m, \overline{K}^m\right] \\
&\underline{K}_m = 0, \overline{K}^m = k, \quad m = 1, 2.
\end{aligned}
$$

Assumption A2 (Payoff shocks). *Fix $\varepsilon > 0$ (for instance, $\varepsilon = 1/2$) and define:*

$$
N_1 = \sigma \quad \text{and} \quad N_2 = (1+\varepsilon)\max(\sigma, \theta, M^c\theta, T\theta)
$$

The payoff process satisfies the following restrictions:

1. *The drift terms are bounded:*
$$\forall t \ge 0, \quad \max(|\nu_t|, |\mu_t|) < N_2.$$

2. *The rate of mean-reversion satisfies:*
$$\int_{t=0}^{+\infty} |\nu_t| dt < N_2.$$

Moreover, the variance is constant and equal to σ, so in particular it is Lipsichtiz and bounded in $[N_1, N_2]$.

Assumption A3 (Strategic complementarities). *For all W and $X > X'$,*

$$D(W,X) - D(W,X') = C(X - X') \ge 0.$$

Assumption A4 (Payoff monotonicity). *For all X and $W > W'$*

$$D(W,X) - D(W',X) = W - W' > 0.\text{[103]}$$

Assumption A5 (Dominance regions). *Let:*

$$
\begin{aligned}
\overline{w} &\equiv -\left(M^c - \frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}(M^c - M^e)\right), \\
\underline{w} &\equiv -\left(M^e + \frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}C\right).
\end{aligned}
\tag{32}
$$

Then,

1. *If $W_t > \overline{w}$, it is strictly dominant for firms in mode 1 to choose $k^1 = 0$ and for firms in mode 2 to choose $k^2 = k$;*

2. *If $W_t < \underline{w}$, it is strictly dominant for firms in mode 1 to choose $k^1 = k$ and for firms in mode 2 to choose $k^2 = 0$.*

[103] In particular, with $w_1 = -\infty$ and $w_2 = +\infty$, the inequality is strict whenever $W, W' \in [w_1, w_2]$.

Proof that Assumption A5 holds. From Lemma 3, we know that if:

$$M_t \leq \underline{\Phi}_t(X_t),$$

then $\tilde{k}(e, M_t, X_t) = 0$ and $\tilde{k}(c, M_t, X_t) = k$ are strictly dominant. Since $\underline{\Phi}_t(.)$ is strictly increasing, a sufficient condition for strict dominance of these choices is therefore:

$$M_t \leq \underline{\Phi}_t(0) = \underline{M}_t$$

For all $t \geq 0$,

$$\underline{M}_t \geq \underline{M}_0 = M_c - \left(\frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}\right)(M^c - M^e).$$

Thus for any $M_t \leq \underline{M}_0$, or equivalently, for any $W_t \geq \overline{w} = -\underline{M}_0$, the choices $\tilde{k}(e, M_t, X_t) = 0$ and $\tilde{k}(c, M_t, X_t) = k$ are strictly dominant. Using $M_t > \overline{\Phi}(X_t)$ as the strict dominance condition for $\tilde{k}(e, M_t, X_t) = k$ and $\tilde{k}(c, M_t, X_t) = 0$, one can similarly obtain the expression reported for $\underline{w}$. □

Assumption A6 (Bounded effects of X on marginal cost)**.** *The derivative of the switching cost functions exist and are equal to zero, and so in particular they satisfy $\partial c_k / \partial X \leq \eta$ for any $\eta \geq 0$.*

Result 1 is then a direct re-statement of Theorems 1 through 7 of Frankel and Burdzy (2005) in the context of our model, so it is left without proof. We can finally use Result 1 to establish Result 2, the threshold property that characterizes the equilibrium dynamics of adoption.

Proof of Result 2. Fix $t \geq 0$ and $X_t \in [0,1]$. By Result 1, the function $M_t \to B_t(M_t, X_t)$ is continuous and strictly decreasing in M_t. Moreover, using Assumption A5, we have that:

$$B_t(\underline{M}, X_t) > 0 \quad \text{and} \quad B_t(\overline{M}, X_t) < 0.$$

By the intermediate value theorem, there exists a unique $\Phi_t(X_t)$ satisfying:

$$B_t(\Phi_t(X_t), X_t) = 0.$$

Moreover, by the implicit function theorem for strictly monotone functions (Dontchev and Rockafellar, 2009, theorem 1H.3), the mapping $(t, X_t) \to \Phi_t(X_t)$ is continuous.

Finally, assume that $C > 0$, and that there exist $X_{1,t} < X_{2,t}$ such that $\Phi_t(X_{1,t}) \geq \Phi_t(X_{2,t})$. Then:

$$\begin{aligned}
0 &= B_t(\Phi_t(X_{1,t}), X_{1,t}) \\
&\leq B_t(\Phi_t(X_{2,t}), X_{1,t}) \\
&< B_t(\Phi_t(X_{2,t}), X_{2,t}) \\
&= 0,
\end{aligned}$$

where the third line uses the strict monotonicity of $B_t(M_t, .)$ when $C > 0$. This is a contradiction; so it must be that $\Phi_t(X_{1,t}) > \Phi_t(X_{2,t})$, and therefore $\Phi_t(.)$ is strictly monotonic.

Continue assuming that $C > 0$. Lemma 3 has established that, if $M_t \geq \overline{\Phi}_t(X_t)$, then $B_t(M_t, X_t) < 0$. In particular, $B_t(\overline{\Phi}_t(X_t), X_t) < 0$, and by strict monotonicity, $\Phi_t(X_t) < \overline{\Phi}_t(X_t)$. The symmetric argument shows that $\underline{\Phi}_t(X_t) < \Phi_t(X_t)$. Finally, the implied dynamics of the adopter share follow from noting that:

$$\tilde{k}(e, M_t, X_t) = k\mathbf{1}\left\{B_t(M_t, X_t) \geq 0\right\} = k\mathbf{1}\left\{M_t \leq \Phi_t(X_t)\right\},$$

and similarly for the case $x_{i,t} = c$.

When $C = 0$, we note that we can solve explicitly for the threshold Φ_t using Lemma 3:

$$\overline{\Phi}_t^{(0)}(X_t) = \underline{\Phi}_t^{(0)}(X_t) = \Phi_t^{(0)} = \underline{M}_t.$$

In particular, for any $C > 0$ and $X_t > 0$, the strict dominance bound for adoption of electronic money, $\underline{\Phi}_t(X_t)$, satisfies:

$$\underline{\Phi}_t(X_t) - \Phi_t^{(0)} = \frac{A_X}{A_{M,t}} X_t > 0,$$

so that the equilibrium adoption threshold $\Phi_t(X_t)$ must also satisfy $\Phi_t(X_t) > \Phi_t^{(0)}$. Since, when $C > 0$, $\Phi_t(0) > \underline{\Phi}_t(0)$, and $\underline{\Phi}_t(0) = \underline{M}_t$, we also have $\Phi_t(0) > \Phi_t^{(0)}(0)$, establishing the result for all $X_t \in [0, 1]$. □

Finally, we note that the proofs of results 1 and 2 make no direct use of the linearity assumption other than to derive strict dominance bounds, so that these results would also hold for more general functional forms for the returns to network scale, so long as this functional form satisfies the assumptions outlined above (and in particular, that it is increasing and Lipschitz-continuous with respect to X and implies strict dominance bounds).

B.2 Proofs for perfect foresight response to large shocks

In this section, we focus on the limit where $T \to +\infty$, and assume that all policy functions are approximately stationary. Moreover, we assume that the shock size satisfies:

$$S > \left(1 + \frac{\theta}{r + k}\right) \frac{M^c - M^e}{M^c}. \tag{33}$$

This condition is sufficient to imply that for any value of C and for any initial user base X_0,

$$M_0 = M^c(1 - S) < \Phi(X_0),$$

so that the economy enters the adoption region at time 0. The assumption $T = +\infty$ can be relaxed, but it is useful to obtain an analytical characterization of the threshold $\underline{C}(X_0)$ above which externalities are sufficiently strong for the shock to lead to permanent adoption. The assumption of perfect foresight also allows for explicit solutions for hitting times. Analog, but weaker results in the general case with shocks are derived in Appendix B.4.

We start by defining perfect foresight trajectories and peak response times formally.

Definition 3 (Perfect foresight trajectory)**.** *Let S satisfying condition (33) and $X_0 \in [0, 1)$. The perfect foresight trajectory of the economy in response to the shock S starting from the user base X_0 is defined as the sample path $\left\{\tilde{M}_t, \tilde{X}_t\right\}_{t \geq 0}$ corresponding to a sequence of innovations to cash demand that are exactly* *equal to zero for all $t > 0$. Such as sequence is given by:*

$$
\begin{aligned}
\tilde{M}_t &= e^{-\theta t} M_0 + (1 - e^{-\theta t}) M^c = (1 - S e^{-\theta t}) M^c, \\
\tilde{X}_t &= e^{-kt} X_0 + \int_0^t e^{-k(t-s)} a(\tilde{M}_s, \tilde{X}_s) ds, \\
a(M_s, X_s) &= \mathbf{1}\left\{ \tilde{M}_s \leq \Phi(\tilde{X}_s) \right\}.
\end{aligned} \tag{34}
$$

Definition 4 (Peak response time)**.** *The peak response time, $\tilde{t}(S, X_0)$, is defined as the first time at which $\tilde{M}_t$ passes through the threshold $\Phi(\tilde{X}_t)$ along the perfect foresight trajectory:*

$$
\hat{t}(X_0) \equiv \inf\left\{ t \geq 0 \mid \tilde{M}_t > \Phi(\tilde{X}_t) \right\}.
$$

Under Condition (33), the peak response time satisfies $\hat{t}(X_0) > 0$.

Next, we characterize adoption dynamics in the perfect foresight case, up to the determination of the peak response time.

Lemma 4 (Adoption dynamics in the perfect foresight case)**.** *Along the perfect foresight trajectory, the adoption decision and the user base are given by:*

$$
\begin{aligned}
a(\tilde{M}_t, \tilde{X}_t) &= \mathbf{1}\left\{ t < \hat{t}(X_0) \right\} \\
\tilde{X}_t &= \begin{cases} 1 - e^{-kt}(1 - X_0) & \text{if} \quad t < \hat{t}(X_0) \\ (1 - e^{-k\hat{t}(X_0)}(1 - X_0)) e^{-k(t - \hat{t}(X_0))} & \text{if} \quad t \geq \hat{t}(X_0) \end{cases}
\end{aligned} \tag{35}
$$

In particular, the user base is increasing for $t < \hat{t}(X_0)$, decreasing for $t \geq \hat{t}(X_0)$, and satisfies:

$$
\lim_{t \to +\infty} \tilde{X}_t = \begin{cases} 0 & \text{if} \quad \hat{t}(X_0) < +\infty \\ 1 & \text{if} \quad \hat{t}(X_0) = +\infty \end{cases} \tag{36}
$$

Proof of Lemma 4. By definition of $\hat{t}(X_0)$, we have that $\forall 0 \leq t < \hat{t}(X_0)$, $\tilde{M}_t \leq \Phi(\tilde{X}_t)$, so that for these values of t, $a(\tilde{M}_t, \tilde{X}_t) = \mathbf{1}\left\{ t < \hat{t}(X_0) \right\}$. If $\hat{t}(X_0) = +\infty$, this is sufficient to establish the expression for the adoption decision. If $\hat{t}(X_0) < +\infty$, let $t \geq \hat{t}(X_0)$ and assume that $\tilde{M}_t > \Phi(\tilde{X}_t)$. In the infinitesimal time period $(t, t + dt)$, the change in the user base is:

$$
d\tilde{X}_t = \left(a(\tilde{M}_t, \tilde{X}_t) - \tilde{X}_t \right) kdt = -\tilde{X}_t kdt < 0.
$$

Thus $\tilde{X}_{t+dt} \leq \tilde{X}_t$. By Result 2, the threshold $\Phi(\tilde{X}_t)$ is increasing; therefore $\Phi(\tilde{X}_{t+dt}) \leq \Phi(\tilde{X}_t)$. Since $d\tilde{M}_t = \theta(M^c - \tilde{M}_t)dt \leq 0$, we have:

$$
\Phi(\tilde{X}_{t+dt}) \leq \Phi(\tilde{X}_t) < \tilde{M}_t \leq \tilde{M}_{t+dt},
$$

implying that $\tilde{M}_{t+dt} > \Phi(\tilde{X}_{t+dt})$ and so $a(\tilde{M}_{t+dt}, \tilde{X}_{t+dt}) = 1$. Thus, for any $t \geq \hat{t}(X_0)$, if $\tilde{M}_t > \Phi(\tilde{X}_t)$, then it must be that $\tilde{M}_{t+dt} > \Phi(\tilde{X}_{t+dt})$. Since the inequality holds at $t = \hat{t}(X_0)$, by induction, $\tilde{M}_t > \Phi(\tilde{X}_t)$ for all $t \geq \hat{t}(X_0)$, and so $a(M_t, X_t) = 1$ at all these dates, giving the expression in the Lemma. Finally, integrating

Equation (34) implies the expressions reported in Equation (35), which in turn implies the limit reported in Equation (36). $\square$

Next, we give the expression for the peak response time when it is finite.

Lemma 5 (Characterization of peak response time)**.** *Let $X_0 \in [0,1)$. Assume that the peak response time is finite: $\hat{t}(X_0) < +\infty$. Then at any date $t \leq \hat{t}(X_0)$, the perfect foresight trajectory satisfies:*

$$\tilde{M}_t = F(\tilde{X}_t), \quad F(x) \equiv M^c \left(1 - S\left(\frac{1-x}{1-X_0}\right)^{\frac{\theta}{k}}\right). \tag{37}$$

Moreover, let $\hat{X}(X_0) = \tilde{X}_{\hat{t}(X_0)}$. Then $\hat{X}(X_0)$ satisfies $F(\hat{X}(X_0)) = \Phi(\hat{X}(X_0))$, and moreover:

$$\hat{t}(X_0) \quad = \frac{1}{k}\log\left(\frac{1-X_0}{1-\hat{X}(X_0)}\right) = \frac{1}{\theta}\log\left(\frac{SM^c}{M^c - \Phi(\hat{X}(X_0))}\right). \tag{38}$$

Proof of Lemma 5. Assume that $\hat{t}(X_0) < +\infty$. Using Lemma 4, the trajectory $\left\{\tilde{M}_t, \tilde{X}_t\right\}_{t \leq \hat{t}(X_0)}$ must satisfy:

$$\tilde{M}_t = (1 - Se^{-\theta t}) \text{ and } \tilde{X}_t = 1 - e^{-kt}(1 - X_0);$$

thus it must satisfy Equation (37). Moreover, if the peak response time is finite, denoting the peak response of the user base by: $\hat{X}(X_0) \equiv \tilde{X}_{\hat{t}(X_0)}$, since the two trajectories $\tilde{M}_t$ and $\tilde{X}_t$ are continuous functions of time, and by result 1, Φ is a continuous function of X_t, the trajectories must satisfy $F(\hat{X}(X_0)) = \Phi(\hat{X}(X_0))$. The expressions for the peak response time follow from $M^c(1 - Se^{-\theta \hat{t}(X_0)}) = \Phi(\hat{X}(X_0)) = F(\hat{X}(X_0))$. $\square$

Finally, we derive sufficient conditions for the peak response time to be either finite or infinite. For this, we use the strict dominance bounds of Lemma 3. These bounds are linear functions of X_t, allowing for a complete characterization of perfect foresight trajectories.

Lemma 6 (Lower bound)**.** *Let $X_0\,[0,1)$. Define:*

$$\underline{\Phi}(X) = M^c - \frac{r+k+\theta}{r+k}(M^c - M^e) + \frac{r+k+\theta}{r+2k}CX.$$

and let $\underline{\hat{t}}(X_0)$ be the peak response time associated with the perfect foresight trajectory generated by $\underline{\Phi}$. Then:

$$\begin{aligned}
&\text{If } \theta = 0, && \underline{\hat{t}}(X_0) \quad = +\infty \\
&\text{If } \theta \in (0,k]\,, && \underline{\hat{t}}(X_0) \quad \begin{cases} < +\infty & \text{if} \quad 0 \leq C < \overline{C}(X_0) \\ = +\infty & \text{if} \quad C \geq \overline{C}(X_0) \end{cases} \\
&\text{If } \theta \in (k,+\infty)\,, && \underline{\hat{t}}(X_0) \quad \begin{cases} < +\infty & \text{if} \quad 0 \leq C \leq \overline{C}(X_0) \\ = +\infty & \text{if} \quad C > \overline{C}(X_0) \end{cases}
\end{aligned} \tag{39}$$

where, when $\theta \in (0,k]$,

$$\overline{C}(X_0) = \frac{r+2k}{r+k}\left(M^c - M^e\right),$$

and when $\theta \in (k, +\infty)$,

$$\overline{C}(X_0) \equiv \frac{r+2k}{r+k+\theta}\frac{\theta}{k}\frac{(1-\overline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}}SM^c,$$

where $\overline{X}(X_0) \in (X_0, 1)$ *is the unique solution to:*

$$\theta(1-X)^{\frac{\theta}{k}-1} - (\theta-k)(1-X)^{\frac{\theta}{k}} = k(1-X_0)^{\frac{\theta}{k}}\frac{r+k+\theta}{r+k}\frac{M^c-M^e}{SM^c}.$$

Finally, when it is finite, the peak response time $\underline{\hat{t}}(X_0)$ *satisfies:*

$$\underline{\hat{t}}(X_0) = \hat{t}^{(0)} + \frac{1}{\theta}\log\left(1 + h(X_0)\right),$$

$$h(X_0) \equiv \frac{(r+k)C\underline{\hat{X}}(X_0)}{(r+2k)(M^c-M^e) - (r+k)C\underline{\hat{X}}(X_0)},$$

where $h(X_0)$ *is positive and strictly increasing when* $C > 0$, *and where* $\hat{t}^{(0)}$ *is the equilibrium peak response time in the model with no external returns,* $C = 0$*:*

$$\hat{t}^{(0)} = \frac{1}{\theta}\log\left(\frac{r+k}{r+k+\theta}\frac{SM^c}{M^c-M^e}\right).$$

Proof of Lemma 6. The peak response time is finite, if and only if, the curve:

$$F(X) = M^c\left(1 - S\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}}\right)$$

intersects the threshold $\underline{\Phi}(X)$ for at least one value $\underline{X}(X_0) \in (X_0, 1)$. Note that $\underline{X}(X_0)$ must be larger than X_0 because of our assumption that:

$$S > \left(1 + \frac{\theta}{r+k}\right)\left(\frac{M^c-M^e}{M^c}\right), \tag{40}$$

and moreover, if the intersection occurs at $X = 1$, then the peak response time is infinite.

If $\theta = 0$, the curve $F(X) = M^c(1-S)$ is flat as a function of X and never intersects $\underline{\Phi}(X)$, given that the assumption in Equation (40) implies that $F(X) < \underline{\Phi} \leq \underline{\Phi}(X)$. So $\hat{t}(X_0) = +\infty$.

If $\theta \in (0, k]$, then the curve $F(X)$ is strictly increasing and either strictly convex (when $\theta < k$), or linear (when $\theta = k$). Note that the assumption in Equation (40) implies that $F(X_0) < \Phi(X_0)$. So a necessary and sufficient condition for $F(X)$ and $\underline{\Phi}(X)$ to intersect on $(X_0, 1)$ is that:

$$\underline{\Phi}(1) < F(1) = M^c,$$

or, after simplifications,

$$C < \frac{r+2k}{r+k}\left(M^c - M^e\right).$$

If $\theta \in (k, +\infty)$, then the curve $F(X)$ is strictly increasing and strictly concave. Let:

$$\Delta(X) \equiv \underline{\Phi}(X) - F(X). \tag{41}$$

The largest value of C for which the two curves $F(X)$ and $\underline{\Phi}(X)$ intersect must be such that the two curves are tangent at their point of intersection; in other words:

$$\underline{\Phi}(\hat{X};C) = F(\hat{X}),$$
$$\frac{\partial \underline{\Phi}(\hat{X};C)}{\partial X} = \frac{\partial F(\hat{X})}{\partial X}.$$

This is equivalent to:

$$\frac{r+k+\theta}{r+2k} C = \frac{SM^c}{1-X_0} \frac{\theta}{k} \left(\frac{1-X}{1-X_0} \right)^{\frac{\theta}{k}-1}, \tag{42}$$

$$\frac{r+k+\theta}{r+k}(M^c - M^e) = \frac{r+k+\theta}{r+2k} CX + SM^c \left(\frac{1-X}{1-X_0} \right)^{\frac{\theta}{k}}. \tag{43}$$

Eliminating C, X must satisfy:

$$g_{LB}(X) = \frac{ak}{S}(1-X_0)^{\frac{\theta}{k}}, \quad g_{LB}(X) = \theta(1-X)^{\frac{\theta}{k}-1} - (\theta-k)(1-X)^{\frac{\theta}{k}}, \quad a \equiv \frac{r+k+\theta}{r+k} \frac{M^c - M^e}{M_c} \tag{44}$$

The right hand side of this equation is strictly larger than 0 and strictly smaller than θ/k, given the assumption that $M^c > M^e$ and the assumption in Equation (40). For any X_0, the function g_{LB} is strictly decreasing on $(X_0, 1)$ and satisfies:

$$g_{LB}(X_0) > k(1-X_0)^{\frac{\theta}{k}}, \quad g_{LB}(1) = 0.$$

Thus Equation (44) has a unique solution $\overline{X}(X_0) \in (X_0, 1)$. Given $\overline{X}(X_0)$, the value of $\overline{C}(X_0)$ is given by:

$$\overline{C}(X_0) = \frac{r+2k}{r+k+\theta} \frac{\theta}{k} \frac{(1-\overline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}} SM^c.$$

For any $C > \overline{C}(X_0)$, the function $\Delta(X)$ has no zero in $(X_0, 1)$. For $C = \overline{C}(X_0)$, it has exactly one zero, which is given by $\overline{X}(X_0)$, which gives the peak response of the adoption trajectory. For $C < \overline{C}(X_0)$, the function $\Delta(X)$ has at least one zero in $\big(X_0, \overline{X}(X_0)\big)$, which also gives the peak response of the adoption trajectory.

Let $\underline{\hat{X}}(X_0)$ be the peak response when $\hat{t}(X_0) < +\infty$. The condition $\Delta(\underline{\hat{X}}(X_0))$ can be written as:

$$e^{-\theta \hat{t}^{(0)}} = e^{-\theta \underline{\hat{t}}(X_0)} + \frac{r+k+\theta}{r+2k} \frac{C}{SM^c} \underline{\hat{X}}(X_0),$$

where the peak response time under $C = 0$ (which is always finite, by the results above), is independent of X_0, and given by:

$$\hat{t}^{(0)} = \frac{1}{\theta} \log \left(\frac{r+k}{r+k+\theta} \frac{SM^c}{M^c - M^e} \right).$$

Manipulating this expression gives the expression for the function $h(.)$. (Note that the peak response time when $C = 0$ is the same for equilibrium adoption trajectory and for the lower bound considered in this lemma, since when $C = 0$, $\underline{\Phi} = \overline{\Phi} = \Phi$, as indicated by Lemma 3). $\square$

Lemma 7 (Upper bound)**.** *Let $X_0\,[0,1)$. Define:*

$$\overline{\Phi}(X_t) = M^c - \frac{r+k+\theta}{r+k}(M^c - M^e) + \frac{r+k+\theta}{r+2k}CX + \frac{r+k+\theta}{r+2k}\frac{k}{r+k}C.$$

and let $\bar{\hat{t}}(X_0)$ be the peak response time associated with the perfect foresight trajectory generated by $\overline{\Phi}$. Then:

$$\begin{array}{lll} \text{If } \theta = 0, & \bar{\hat{t}}(X_0) & = +\infty \\ \text{If } \theta \in (0,k]\,, & \bar{\hat{t}}(X_0) & \begin{cases} < +\infty & \text{if} \quad 0 \le C < \underline{C}(X_0) \\ = +\infty & \text{if} \quad C \ge \underline{C}(X_0) \end{cases} \\ \text{If } \theta \in (k,+\infty)\,, & \bar{\hat{t}}(X_0) & \begin{cases} < +\infty & \text{if} \quad 0 \le C \le \underline{C}(X_0) \\ = +\infty & \text{if} \quad C > \underline{C}(X_0) \end{cases} \end{array} \tag{45}$$

where, when $\theta \in (0,1]$,

$$\underline{C}(X_0) = M^c - M^e,$$

and when $\theta \in (k,+\infty)$,

$$\underline{C}(X_0) \equiv \frac{r+2k}{r+k+\theta}\frac{\theta}{k}\frac{(1-\underline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}}SM^c.$$

where $\underline{X}(X_0) \in (X_0, 1)$ is the unique solution to:

$$\left(1+\frac{k}{r+k}\right)\theta(1-X)^{\frac{\theta}{k}-1} - (\theta-k)(1-X)^{\frac{\theta}{k}} = k(1-X_0)^{\frac{\theta}{k}}\frac{r+k+\theta}{r+k}\frac{M^c-M^e}{SM^c}.$$

Proof of Lemma 7. Note that the assumption that:

$$S > \left(1+\frac{\theta}{r+k}\right)\left(\frac{M^c-M^e}{M^c}\right), \tag{46}$$

implies that $\overline{\Phi}(X_0) > F(X_0)$, as for the case of the lower threshold $\underline{\Phi}$.

The case $\theta \in [0,k]$ is similar to the proof of Lemma 6, noting that the condition $\overline{\Phi}(1) < F(1)$ is equivalent to:

$$C < M^c - M^e.$$

If $\theta \in (k,+\infty)$, the conditions characterizing the largest value of C for which there is an intersection between the two curves becomes:

$$\frac{r+k+\theta}{r+2k}C = \frac{SM^c}{1-X_0}\frac{\theta}{k}\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}-1}, \tag{47}$$

$$\frac{r+k+\theta}{r+k}\left(M^c - M^e - \frac{k}{r+2k}C\right) = \frac{r+k+\theta}{r+2k}CX + SM^c\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}}. \tag{48}$$

Eliminating C, X must satisfy:

$$g_{UB}(X) = \frac{ak}{S}(1-X_0)^{\frac{\theta}{k}}, \quad g_{UB}(X) \equiv \left(1+\frac{k}{r+k}\right)\theta(1-X)^{\frac{\theta}{k}-1}-(\theta-k)(1-X)^{\frac{\theta}{k}}, \quad a \equiv \frac{r+k+\theta}{r+k}\frac{M^c-M_e}{M_c}. \tag{49}$$

The right hand side of this equation is strictly larger than 0 and strictly smaller than θ/k, given the assumption that $M^c > M^e$ and the assumption in Equation (40). For any X_0, the function g_{UB} is strictly decreasing on $(X_0, 1)$ and satisfies:

$$g_{UB}(X_0) > k(1-X_0)^{\frac{\theta}{k}}, \quad g_{UB}(1) = 0.$$

Thus Equation (49) has a unique solution $\underline{X}(X_0) \in (X_0, 1)$. Given $\underline{X}(X_0)$, the value of $\underline{C}(X_0)$ is given by:

$$\underline{C}(X_0) = \frac{r+2k}{r+k+\theta}\frac{\theta}{k}\frac{(1-\underline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}}SM^c.$$

For any $C > \underline{C}(X_0)$, the function $\Delta(X) = F(X) - \overline{\Phi}(X)$ has no zero in $(X_0, 1)$. For $C = \underline{C}(X_0)$, it has exactly one zero, which is given by $\underline{X}(X_0)$, which gives the peak response of the adoption trajectory. For $\underline{C}(X_0)$, the function $\Delta(X)$ has at least one zero in $(X_0, \underline{X}(X_0))$, which also gives the peak response of the adoption trajectory. $\square$

Lemma 8 (Monotonicity)**.** *Consider two perfect foresight trajectories, $\left\{\tilde{M}_t, \tilde{X}_t^{(a)}\right\}$ and $\left\{\tilde{M}_t, \tilde{X}_t^{(b)}\right\}$, associated with the same shock S and the same initial condition X_0, but generated by two adoption thresholds $\Phi_t^{(a)}$ and $\Phi_t^{(b)}$ that satisfy:*

$$\forall X \in [0,1], \quad \forall t \geq 0, \quad \Phi_t^{(a)}(X) \leq \Phi_t^{(b)}(X).$$

Then the two trajectories satisfy:

$$\hat{t}_{(a)}(S, X_0) \leq \hat{t}_{(b)}(S, X_0) \quad \textit{and} \quad \forall t \geq 0, \quad \tilde{X}_t^{(a)} \leq \tilde{X}_t^{(b)} \quad .$$

Proof of Lemma 8. Let $t \geq 0$ and assume that $\tilde{X}_t^{(a)} \leq \tilde{X}_t^{(b)}$. In the infinitesimal time interval $(t, t+dt)$, using the law of motion for each of the perfect foresight trajectories,

$$\tilde{X}_{t+dt}^{(b)} - \tilde{X}_{t+dt}^{(a)} = (\tilde{X}_t^{(b)} - \tilde{X}_t^{(a)})(1-kdt) + \left(\mathbf{1}\left\{\tilde{M}_t \leq \Phi_t^{(b)}(\tilde{X}_t^b)\right\} - \mathbf{1}\left\{\tilde{M}_t \leq \Phi_t^{(a)}(\tilde{X}_t^a)\right\}\right)kdt. \tag{50}$$

Since $\Phi_t^{(b)}(\tilde{X}_t^{(b)}) \geq \Phi_t^{(b)}(\tilde{X}_t^{(a)}) \geq \Phi_t^{(a)}(\tilde{X}_t^{(a)})$,

$$\mathbf{1}\left\{\tilde{M}_t \leq \Phi_t^{(b)}(\tilde{X}_t^b)\right\} \geq \mathbf{1}\left\{\tilde{M}_t \leq \Phi_t^{(a)}(\tilde{X}_t^a)\right\}.$$

Therefore, $\tilde{X}_t^b \geq \tilde{X}_t^a$ implies $\tilde{X}_{t+dt}^{(b)} \geq \tilde{X}_{t+dt}^{(a)}$. Since $\tilde{\tilde{X}}_0^{(b)} = \tilde{\tilde{X}}_0^{(a)} = X_0$, by induction, $\tilde{X}_t^{(a)} \leq \tilde{X}_t^{(b)}$. Therefore, along the two trajectories, we have $\Phi_t^{(b)}(X_t^{(b)}) \geq \Phi_t^{(a)}(X_t^{(a)})$, so that

$$\forall t \geq 0, \tilde{M}_t > \Phi_t^{(b)}(X_t^{(b)}) \implies \tilde{M}_t > \Phi_t^{(a)}(X_t^{(a)}),$$

implying that $\hat{t}_{(a)}(S, X_0) \leq \hat{t}_{(b)}(S, X_0)$. $\square$

Lemma 9 (Bounds on equilibrium peak response times)**.** *Let $\hat{t}(X_0)$ be the peak response times associated with the equilibrium threshold Φ. Then,*

$$\forall (S, X_0), \quad \underline{\hat{t}}(X_0) \leq \hat{t}(X_0) \leq \overline{\hat{t}}(X_0).$$

Proof of Lemma 9. By Lemma 3, the equilibrium threshold and the two strict dominance thresholds satisfy: $\underline{\Phi}(X) \leq \Phi(X) \leq \overline{\Phi}(X)$. Applying Lemma 8 then gives the result. □

This completes the characterization of the equilibrium adoption dynamics in perfect foresight. In particular, the partition of the space of (C, X_0) into regions corresponding to finite and infinite peak response times, and reported in Figure 3 and Appendix Figure H.16 follow from Lemmas 6, 7 and 9. We conclude by mapping these results to Predictions 1a, 2a, and 3a in Section 3.2.1.

Proof of Predictions 1a and 2a. By Result 2, when $C > 0$, the equilibrium adoption threshold satisfies $\Phi(X) > \Phi^{(0)}(X)$, where $\Phi^{(0)}(X)$ is the adoption threshold when $C = 0$. By Lemma 8, this implies that $\hat{t}(X_0) > \hat{t}^{(0)}$. Moreover, lemma 9 indicates that $\hat{t}(X_0) \geq \underline{\hat{t}}(X_0)$. When $C > \overline{C}(X_0)$, by Lemma 6, $\underline{\hat{t}}(X_0) = +\infty$, implying that $\hat{t}(X_0) = +\infty$. □

Note that Equation (14), in the main text, follows from Lemmas 6 and 8.

Proof of Prediction 3a. By Lemma 6, when the peak response time exists, it must satisfy:

$$F(\hat{X}(X_0); X_0) = \Phi(\hat{X}(X_0)).$$

Applying the implicit function theorem, we obtain:

$$\frac{\partial \hat{X}}{\partial X_0} = \frac{SM^c(1-X)^{\frac{\theta}{k}-1}(1-X_0)^{-\frac{\theta}{k}}\frac{\theta}{k}}{SM^c(1-X)^{\frac{\theta}{k}-1}(1-X_0)^{-\frac{\theta}{k}-1}\frac{\theta}{k} + \Phi'(\hat{X}(X_0))} > 0. \tag{51}$$

Thus $\hat{X}(X_0)$ increases with X_0. By Lemma 5, when it is finite, the peak response time satisfies:

$$\hat{t}(X_0) = \frac{1}{\theta} \log\left(\frac{SM^c}{M^c - \Phi(\hat{X}(X_0))}\right). \tag{52}$$

The right-hand side is an increasing function of $\hat{X}(X_0)$, yielding the result. □

B.3 Perfect foresight response when $T < +\infty$

In this section, we analyze the perfect foresight response of the economy when $T < +\infty$, so that, contrary to Section B.2, policy functions may not be stationary. We make the following two assumptions:

$$S > \left(1 + \frac{\theta}{r+k}\right)\frac{M^c - M^e}{M^c}, \tag{53}$$

$$T > \frac{1}{\theta}\log\left(\frac{SM^c}{M^c - M^e}\right). \tag{54}$$

The first condition is identical to condition (33) and ensures that there is adoption on impact, even when $C = 0$. As explained below, the second condition ensures that in the case where $C = 0$, the perfect foresight response has a finite peak response time. Peak response times and perfect foresight trajectories are defined similarly to B.2.

Definition 5 (Perfect foresight trajectory). *Let S satisfying condition (53) and $X_0 \in [0,1)$. The perfect foresight trajectory of the economy associated with an arbitrary sequence of thresholds $\{\Phi_t\}_{t\geq 0}$ in response to the shock S starting from the user base X_0 is defined as the sample path $\left\{\tilde{M}_t, \tilde{X}_t\right\}_{t\geq 0}$ corresponding to a sequence of innovations to cash demand that are exactly equal to zero for all $t > 0$. Such as sequence is given by:*

$$\begin{aligned}
\tilde{M}_t &= \begin{cases} (1 - Se^{-\theta t})M^c & \text{if} \quad 0 \leq t \leq T \\ (1 - Se^{-\theta T})M^c & \text{if} \quad t > T \end{cases} \\
\tilde{X}_t &= e^{-kt} X_0 + \int_0^t e^{-k(t-s)} a_t(\tilde{M}_s, \tilde{X}_s) ds, \\
a_t(M_s, X_s) &= \mathbf{1}\left\{ \tilde{M}_s \leq \Phi_t(\tilde{X}_s) \right\}.
\end{aligned} \tag{55}$$

Definition 6 (Peak response time). *The peak response time, $\tilde{t}(S, X_0)$, is defined as the first time at which $\tilde{M}_t$ passes through the threshold $\Phi_t(\tilde{X}_t)$ along the perfect foresight trajectory:*

$$\hat{t}(X_0) \equiv \inf\left\{ t \geq 0 \mid \tilde{M}_t > \Phi_t(\tilde{X}_t) \right\}.$$

Under Condition 53, the peak response time satisfies $\hat{t}(X_0) > 0$.

We establish the three following predictions in the case $T < +\infty$.

Prediction 1d. ***(Persistent response of the user base)*** *When $C > 0$, the response of the user base satisfies $\tilde{X}_t \geq \tilde{X}_t^{(0)}$ for all $t \geq 0$, where $\tilde{X}^{(0)}$ is the perfect foresight adoption trajectory when $C = 0$. Moreover, when $C > \overline{\overline{C}}(X_0)$, $\lim\limits_{t\to+\infty} \tilde{X}_t = 1 > X_0$, where the expression for $\hat{C}(X_0)$ is given in Lemma (12).*

Prediction 2d. ***(Persistent response of the adoption rate)*** *When $C > 0$, the adoption response is $a_t = 1$ for all $t \leq \hat{t}(X_0)$, where the peak response time satisfies $\hat{t}(X_0) \geq \hat{t}^{(0)}$. Moreover, when $C \geq \overline{\overline{C}}(X_0)$, $\hat{t}(X_0) = +\infty$.*

Prediction 3d. ***(Positive state-dependence with respect to the initial user base)*** *When $C > 0$, the persistence of the response satisfies:*

$$\hat{t}(X_0) \geq \underline{\underline{t}}(X_0),$$

where the function $\underline{\underline{t}}(X_0)$ is increasing with X_0.

Relative to the case $T \to +\infty$, main intuitions from Predictions 1a-3a are preserved, but there are two main differences. The first is that we do not provide a general partition of the long-run behavior of the user share as a function of the initial adoption rate, X_0, and the strength of complementarities, C, as in Figure 3. The second is that Prediction 3a is slightly weaker: we cannot establish that the peak response time itself is increasing with the user base, but only that it is bounded from below by a function that is increasing with respect to the user base.

We start by characterizing the perfect foresight trajectory.

Lemma 10 (Characterization of perfect foresight trajectory). *Let $X_0 \in [0,1)$. Let $\{\Phi_t\}_{t\geq 0}$ be an arbitrary sequence of positive, increasing thresholds satisfying: $\Phi_t = \Phi_T$ for all $t \geq T$. Assume that the peak response* *time associated with the sequence of thresholds is finite:* $\hat{t}(X_0) < +\infty$*. Then it must be smaller than* T*:* $\hat{t}(X_0) \leq T$*. Moreover, at any date* $t \leq \hat{t}(X_0)$*, the perfect foresight trajectory satisfies:*

$$\tilde{M}_t = F(\tilde{X}_t), \quad F(x) \equiv M^c \left(1 - S\left(\frac{1-x}{1-X_0}\right)^{\frac{\theta}{k}}\right). \tag{56}$$

Moreover, let $\hat{X}(X_0) = \tilde{X}_{\hat{t}(X_0)}$*. Then* $\hat{X}(X_0)$ *satisfies* $F(\hat{X}(X_0)) = \Phi_{\hat{t}(X_0)}(\hat{X}(X_0))$*, and moreover:*

$$\hat{t}(X_0) \quad = \frac{1}{k}\log\left(\frac{1-X_0}{1-\hat{X}(X_0)}\right) = \frac{1}{\theta}\log\left(\frac{SM^c}{M^c - \Phi_{\hat{t}(X_0)}(\hat{X}(X_0))}\right). \tag{57}$$

This result holds in particular when the sequence of thresholds is the equilibrium sequence characterized in Result 1. Therefore, when the peak response time of the perfect foresight trajectory associated with the equilibrium sequence of thresholds is finite, it must be smaller than T.

Proof of Lemma 10. Assume that $\hat{t}(X_0) < +\infty$. First, note that by definition of the peak response time, for any $t \leq \hat{t}(X_0)$, we must have:

$$X_t = 1 - e^{-kt}(1 - X_0).$$

Next, note that if it is finite, the the peak response time must satisfy $\hat{t}(X_0) \leq T$. Assume otherwise, that is, $+\infty > \hat{t}(X_0) > T$. Then it must be that:

$$\Phi_T(1 - e^{-kT}(1 - X_0)) > M_T = (1 - Se^{-\theta T})M^c$$

Since $\Phi_t = \Phi_T$ for all $t \geq T$, and since mean-reversion vanishes for $t \geq T$, we must have:

$$\Phi_t(1 - e^{-kt}(1 - X_0)) = \Phi_T(1 - e^{-kt}(1 - X_0)) \geq \Phi_T(1 - e^{-kT}(1 - X_0)) > M_T = M_t$$

so that there is no adoption for $t \geq T$. Therefore, $\hat{t}(X_0) = +\infty$, a contradiction. Thus when it is finite, the peak response time must satisfy $\hat{t}(X_0) \leq T$.

Using Definition 55, the trajectory $\left\{\tilde{M}_t, \tilde{X}_t\right\}_{t \leq \hat{t}(X_0)}$ must satisfy:

$$\tilde{M}_t = (1 - Se^{-\theta t}) \text{ and } \tilde{X}_t = 1 - e^{-kt}(1 - X_0),$$

and therefore Equation (56). Moreover, if the peak response time is finite, denoting the peak response of the user base by: $\hat{X}(X_0) \equiv \tilde{X}_{\hat{t}(X_0)}$, since the two trajectories $\tilde{M}_t$ and $\tilde{X}_t$ are continuous functions of time, and since by result 1, Φ is a continuous function of X_t, the trajectories must satisfy $F(\hat{X}(X_0)) = \Phi_{\hat{t}(X_0)}(\hat{X}(X_0))$. The expressions for the peak response time follow from $M^c(1-Se^{-\theta \hat{t}(X_0)}) = \Phi_{\hat{t}(X_0)}(\hat{X}(X_0)) = F(\hat{X}(X_0))$. ☐

Lemma 11 (The perfect foresight trajectory when $C = 0$)**.** *Assume that* $C = 0$*. Then if condition 54 holds, the peak response time when* $C = 0$ *is the unique solution to:*

$$\hat{t}^{(0)} = \frac{1}{\theta}\log\left(\frac{SM^c}{M^c - M^e}\frac{r + k + \theta e^{-(r+k+\theta)(T-\hat{t}^{(0)})}}{r + k + \theta}\right). \tag{58}$$

Proof of Lemma 11. In this case, using Lemma 3, the equilibrium sequence of thresholds is independent of

X_t, and given by:

$$\underline{\Phi}_t^{(0)} = \begin{cases} M^c - \dfrac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)(T-t)}}(M^c - M^e) & \text{if} \quad t \leq T \\ M^e & \text{if} \quad t \geq T \end{cases}$$

Using this expression along with Lemma B.3, if the peak response time is finite, it must satisfy:

$$\hat{t}^{(0)} = \frac{1}{\theta} \log \left(\frac{SM^c}{M^c - M^e} \frac{r+k+\theta e^{-(r+k+\theta)(T-\hat{t}^{(0)})}}{r+k+\theta} \right) \tag{59}$$

Denoting:

$$v(t) = t - \frac{1}{\theta} \log \left(\frac{SM^c}{M^c - M^e} \frac{r+k+\theta e^{-(r+k+\theta)(T-t)}}{r+k+\theta} \right), \tag{60}$$

computation shows that $v'(t) > 0$. Moreover, Assumption 54 implies $v(T) > 0$, while Assumption 53 implies $v(0) < 0$. So there is a unique solution, $\hat{t}^{(0)}$, to Equation (59). Moreover, it is straightforward to check that, because of the monotonicity of $v(t)$, for all $t < \hat{t}^{(0)}$, $\tilde{M}_t < \underline{\Phi}_t$, while for all $T \geq t \geq \hat{t}^{(0)}$, $\tilde{M}_t \geq \underline{\Phi}_t$. Thus the peak response time is finite and equal to $\hat{t}^{(0)}$. □

Lemma 3 shows that a lower bound on equilibrium thresholds is given by:

$$\underline{\Phi}_t(X) = \begin{cases} M^c + \dfrac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)(T-t)}} \left(\dfrac{r+k}{r+2k} CX - (M^c - M^e) \right) & \text{if} \quad t \leq T \\ M^e + \dfrac{r+k}{r+2k} CX & \text{if} \quad t \geq T \end{cases}$$

Define the sequence of (time-invariant) thresholds $\left\{ \underline{\underline{\Phi}}_t \right\}_{t \geq 0}$ by:

$$\underline{\underline{\Phi}}_t(X) = I_T + \frac{r+k}{r+2k} CX, \quad I_T \equiv M^c - \frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}} (M^c - M^e).$$

Computation shows that:

$$\forall t \geq 0, \forall X \in [0,1], \quad \underline{\Phi}_t(X) \geq \underline{\underline{\Phi}}_t(X). \tag{61}$$

We next prove the following lemma regarding the perfect foresight trajectory generated by the sequence $\left\{ \underline{\underline{\Phi}}_t \right\}_{t \geq 0}$, which is an analog of Lemma (6).

Lemma 12 (Lower bound)**.** *Let $\hat{\underline{\underline{t}}}(X_0)$ be the peak response time associated with the perfect foresight trajectory generated by the sequence $\left\{ \underline{\underline{\Phi}}_t \right\}_{t \geq 0}$. Then:*

$$\begin{aligned} &\textit{If } \theta = 0, && \hat{\underline{\underline{t}}}(X_0) \quad = +\infty \\ &\textit{If } \theta \in (0, k], && \hat{\underline{\underline{t}}}(X_0) \quad \begin{cases} < +\infty & \textit{if} \quad 0 \leq C < \overline{\overline{C}}(X_0) \\ = +\infty & \textit{if} \quad C \geq \overline{\overline{C}}(X_0) \end{cases} \\ &\textit{If } \theta \in (k, +\infty), && \hat{\underline{\underline{t}}}(X_0) \quad \begin{cases} < +\infty & \textit{if} \quad 0 \leq C \leq \overline{\overline{C}}(X_0) \\ = +\infty & \textit{if} \quad C > \overline{\overline{C}}(X_0) \end{cases} \end{aligned}$$

where, when $\theta \in (0, k]$,

$$\overline{\overline{C}}(X_0) = \frac{r+2k}{(r+k)(1-e^{-kT}(1-X_0))}\left(\frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}(M^c - M^e) - Se^{-\theta T}M^c\right),$$

and when $\theta \in (k, +\infty)$,

$$\overline{\overline{C}}(X_0) \equiv \frac{r+2k}{r+k}\frac{\theta}{k}\frac{(1-\overline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}}SM^c,$$

where $\overline{X}(X_0) \in (X_0, 1)$ is the unique solution to:

$$\theta(1-X)^{\frac{\theta}{k}-1} - (\theta - k)(1-X)^{\frac{\theta}{k}} = k(1-X_0)^{\frac{\theta}{k}}\frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}\frac{M^c - M^e}{SM^c}.$$

Proof of Lemma 12. The proof is similar to the proof Lemma 6, but for a different threshold. Where there is no difference between proofs, we refer to the that proof.

The peak response time is finite, if and only if, the curve:

$$F(X) = M^c\left(1 - S\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}}\right)$$

intersects the threshold $\underline{\underline{\Phi}}(X)$ for at least one value $\underline{\underline{X}}(X_0) \in \big(X_0, 1-e^{-kT}(1-X_0)\big)$. Note that $\underline{\underline{X}}(X_0)$ must be larger than X_0 because of Assumption (53), and it must be smaller than $1-e^{-kT}(1-X_0)$ because by Lemma (11), if the peak response time associated with a threshold is finite, it must be smaller than T.

If $\theta = 0$, the proof the same as for Lemma 6.

If $\theta \in (0, k]$, following similar arguments as in the proof of Lemma 6, a necessary and sufficient condition for $F(X)$ and $\underline{\underline{\Phi}}(X)$ to intersect on $\big(X_0, 1-e^{-kT}(1-X_0)\big)$ is that:

$$\underline{\underline{\Phi}}(1-e^{-kT}(1-X_0)) < F(1-e^{-kT}(1-X_0)),$$

or, after simplifications,

$$C < \frac{r+2k}{(r+k)(1-e^{-kT}(1-X_0))}\left(\frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}(M^c - M^e) - Se^{-\theta T}M^c\right).$$

If $\theta \in (k, +\infty)$, then the curve $F(X)$ is strictly increasing and strictly concave. The largest value of C for which the two curves $F(X)$ and $\underline{\underline{\Phi}}(X)$ intersect must be such that the two curves are tangent at their point of intersection; in other words:

$$\underline{\underline{\Phi}}(\hat{X}) = F(\hat{X}), \quad \frac{\partial\underline{\underline{\Phi}}(\hat{X})}{\partial X} = \frac{\partial F(\hat{X})}{\partial X}.$$

This is equivalent to:

$$\frac{r+k}{r+2k}C = \frac{SM^c}{1-X_0}\frac{\theta}{k}\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}-1}, \tag{62}$$

$$\frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}(M^c - M^e) = \frac{r+k}{r+2k}CX + SM^c\left(\frac{1-X}{1-X_0}\right)^{\frac{\theta}{k}}. \tag{63}$$

Eliminating C, X must satisfy:

$$g_{LB}(X) = \frac{ak}{S}(1-X_0)^{\frac{\theta}{k}}, \quad g_{LB}(X) \equiv \theta(1-X)^{\frac{\theta}{k}-1} - (\theta-k)(1-X)^{\frac{\theta}{k}}, \quad a \equiv \frac{r+k+\theta}{r+k+\theta e^{-(r+k+\theta)T}}\frac{M^c - M^e}{M_c} \tag{64}$$

Note that $a/S < 1$. The right hand side of this equation is strictly larger than 0 and strictly smaller than θ/k, given the assumption that $M^c > M^e$ and the assumption in Equation (53). For any X_0, the function g_{LB} is strictly decreasing on $(X_0, 1)$ and satisfies:

$$g_{LB}(X_0) > k(1-X_0)^{\frac{\theta}{k}} > \frac{ak}{S}(1-X_0)^{\frac{\theta}{k}}, \quad g_{LB}(1-e^{-kT}(1-X_0)) < \frac{ak}{S}(1-X_0)^{\frac{\theta}{k}}.$$

Thus Equation (64) has a unique solution $\overline{X}(X_0) \in \big(X_0, 1-e^{-kT}(1-X_0)\big)$. Given $\overline{\overline{X}}(X_0)$, the value of $\overline{\overline{C}}(X_0)$ is given by:

$$\overline{\overline{C}}(X_0) = \frac{r+2k}{r+k}\frac{\theta}{k}\frac{(1-\overline{X}(X_0))^{\frac{\theta}{k}-1}}{(1-X_0)^{\frac{\theta}{k}}}SM^c.$$

For any $C > \overline{\overline{C}}(X_0)$, the function $\Delta(X) \equiv F(X) - \underline{\underline{\Phi}}(X)$ has no zero in $\big(X_0, 1-e^{-kT}(1-X_0)\big)$. For $C = \overline{\overline{C}}(X_0)$, it has exactly one zero, which is given by $\overline{\overline{X}}(X_0)$, which gives the peak response of the adoption trajectory. For $\overline{\overline{C}}(X_0) > C$, the function $\Delta(X)$ has at least one zero in $\Big(X_0, \overline{\overline{X}}(X_0)\Big)$, which also gives the peak response of the adoption trajectory.

□

We are now in a position to prove Predictions 1d, 2d and 3d.

Proof of Predictions 1d and 2d. Fix $C > 0$ and $\left\{\Phi_t^{(0)}\right\}_{t\geq 0}$ be the corresponding equilibrium sequence of adoption thresholds and $\tilde{X}_t$ the perfect foresight adoption trajectory. First, note that by Lemma 3, the equilibrium sequence of adoption thresholds $\left\{\Phi_t^{(0)}\right\}_{t\geq 0}$ for the case $C = 0$ satisfy $\Phi_t \geq \Phi_t^{(0)}$. By Lemma 8, the perfect foresight trajectory must therefore satisfy $\tilde{X}_t \geq \tilde{X}_t^{(0)}$.

Let $\left\{\underline{\underline{\tilde{X}}}_t\right\}_{t\geq 0}$ be the perfect foresight trajectory associated with the (time-invariant) sequence of thresholds $\left\{\underline{\underline{\Phi}}_t\right\}_{t\geq 0}$. Since $\Phi_t \geq \underline{\Phi}_t \geq \underline{\underline{\Phi}}_t$ for all $t \geq 0$, we have that $\tilde{X}_t \geq \underline{\underline{\tilde{X}}}_t$ and $\hat{t}(X_0) \geq \underline{\underline{\hat{t}}}(X_0)$. Lemma 12 then establishes the rest of the results in Predictions 1d and 2d. □

Proof of Prediction 3d. The proof of Predictions 1d and 2d establishes that

$$\hat{t}(X_0) \geq \underline{\underline{\hat{t}}}(X_0).$$

Thus what remains to be established is that $\underline{\underline{\hat{t}}}(X_0)$ is increasing with respect to X_0. Let $\underline{\underline{\hat{X}}}(X_0)$ denote the peak response of the user base under the (time-invariant) sequence of adoption thresholds $\left\{\underline{\underline{\Phi}}_t\right\}_{t\geq 0}$. Following the same steps as in the proof of Prediction 3d, we can show that $\underline{\underline{\hat{X}}}(X_0)$ is increasing with X_0. Lemma 10 then shows that:

$$\hat{t}(X_0) = \frac{1}{\theta}\log\left(\frac{SM^c}{M^c - \underline{\underline{\Phi}}(\underline{\underline{\hat{X}}}(X_0))}\right). \tag{65}$$

The right-hand side is an increasing function of $\underline{\underline{\hat{X}}}(X_0)$, yielding the result. $\square$

B.4 Response to large shocks: general case

Next, we prove Predictions 1b and 2b. To do this, we first prove Lemma 13, which states the following. Define the partial ordering $\succ$ on sequences of adoption thresholds (that is, sequences of continuous, increasing and real-valued functions over $[0,1]$) by:

$$\Phi^{(2)} \succ \Phi^{(1)} \iff \forall t \geq 0, \forall X \in [0,1]\,, \Phi_t^{(2)}(X) > \Phi_t^{(1)}(X).$$

Lemma 13 states that when two sequences of adoption thresholds satisfy $\Phi^{(2)} \succ \Phi^{(1)}$, then their IRFs inherit the ordering, that is, the IRFs associated with $\Phi^{(2)}$ are strictly higher than those associated with $\Phi^{(1)}$.

Lemma 13 (Threshold monotonicity)**.** *Consider two sequences of adoption thresholds satisfying $\Phi^{(2)} \succ \Phi^{(1)}$. For $i = 1, 2$, define:*

$$dX_t^{(i)} = \left(a_t^{(i)}(M_t, X_t^{(i)}) - X_t^{(i)}\right) kdt$$

$$a_t^{(i)}(M_t, X_t^{(i)}) = \mathbf{1}\left\{M_t \leq \Phi_t^{(i)}\left(X_t^{(i)}\right)\right\}$$

where M_t is the stochastic process defined in Equation (3). Then, for any $t > 0$, $S > 0$ and $X \in [0, 1)$,

$$\begin{aligned}\mathcal{I}_a^{(1)}(t, S, X) &\equiv \mathbb{E}_0\left[a_t\left(M_t, X_t^{(1)}\right) \mid M_0 = (1-S)M^c, X_0^{(1)} = X\right] \\ &< \mathbb{E}_0\left[a_t\left(M_t, X_t^{(2)}\right) \mid M_0 = (1-S)M^c, X_0^{(2)} = X\right] \equiv \mathcal{I}_a^{(2)}(t, S, X)\,.\end{aligned}$$

and

$$\begin{aligned}\mathcal{I}_X^{(1)}(t, S, X) &\equiv \mathbb{E}_0\left[X_t\left(M_t, X_t^{(1)}\right) \mid M_0 = (1-S)M^c, X_0^{(1)} = X\right] \\ &< \mathbb{E}_0\left[X_t\left(M_t, X_t^{(2)}\right) \mid M_0 = (1-S)M^c, X_0^{(2)} = X\right] \equiv \mathcal{I}_X^{(2)}(t, S, X)\,.\end{aligned}$$

Proof of Lemma 13. Fix initial conditions $(M_0 = M^c(1-S), X_0 = X)$, and fix a particular sample path for cash-based demand, $M = \{\tilde{M}_t\}_{t\geq 0}$, that is, the values cash-based demand starting associated with the initial condition M_0 and a particular sequence of exogenous innovations $\{d\tilde{Z}_t\}_{t\geq 0}$. Let the user base $\{X_t^{(M,i)}\}$ generated by the sample path M using the adoption threshold $\Phi^{(i)}$ be defined as:

$$X_t^{(M,i)} = k\int_0^t a_s^{(i)}(M_s, X_s^{(M,i)})ds + e^{-kt}X, \quad i = 1, 2.$$

We next show by induction that $X_t^{(M,1)} \leq X_t^{(M,2)}$ for all $t \geq 0$. Consider a particular date t and assume that $X_t^{(M,1)} \leq X_t^{(M,2)}$. By assumption, $\Phi_t^{(1)}(X) < \Phi_t^{(2)}(X)$ for any $X \in [0,1]$, and both thresholds are increasing, so:

$$\Phi_t^{(1)}(X_t^{(1)}) \leq \Phi_t^{(1)}(X_t^{(2)}) < \Phi_t^{(2)}(X_t^{(2)}).$$

Therefore, $a_t^{(1)}(M_t, X_t^{(M,i)}) \leq a_t^{(2)}(M_t, X_t^{(M,i)})$. In an infinitesimal time period, we have, for $i = 1, 2$:

$$X_{t+dt}^{(M,i)} = (1-kdt)X_t^{(M,i)} + a_t^{(i)}(M_t, X_t^{(M,i)})kdt.$$

When $X_t^{(M,1)} \leq X_t^{(M,2)}$, this implies that $X_{t+dt}^{(M,1)} \leq X_{t+dt}^{(M,2)}$. Since by assumption $X_0^{(M,1)} = X_0^{(M,2)} = X$, this proves that $X_t^{(M,1)} \leq X_t^{(M,2)}$ for all $t \geq 0$. In turn, along the sample path M, we therefore have $\Phi_t^{(1)}(X_t^{(M,1)}) < \Phi_t^{(2)}(X_t^{(M,2)})$, where the inequality is strict because of the assumption that $\Phi_t^{(1)}(X) < \Phi_t^{(2)}(X)$ for all $X \in [0,1]$. Therefore:

$$\mathbf{1}\left\{M_t \leq \Phi_t^{(2)}\left(X_t^{(M,2)}\right)\right\} - \mathbf{1}\left\{M_t \leq \Phi_t^{(1)}\left(X_t^{(M,1)}\right)\right\} = \mathbf{1}\left\{\Phi_t^{(1)}\left(X_t^{(M,1)}\right) < M_t \leq \Phi_t^{(2)}\left(X_t^{(M,2)}\right)\right\}, \quad (66)$$

where the interval $\left(\Phi_t^{(1)}\left(X_t^{(M,1)}\right), \Phi_t^{(2)}\left(X_t^{(M,2)}\right)\right]$ has a non-empty interior. Then, note that:

$$\begin{aligned}
\mathcal{I}_a^{(2)}(t,S,X) - \mathcal{I}_a^{(1)}(t,S,X) &= \mathbb{E}_0\left[\mathbf{1}\left\{M_t \leq \Phi_t^{(2)}\left(X_t^{(2)}\right)\right\} \mid M_0, X_0\right] - \mathbb{E}_0\left[\mathbf{1}\left\{M_t \leq \Phi_t^{(1)}\left(X_t^{(1)}\right)\right\} \mid M_0, X_0\right] \\
&= \mathbb{E}_0\left[\mathbf{1}\left\{\Phi_t^{(1)}\left(X_t^{(1)}\right) < M_t \leq \Phi_t^{(2)}\left(X_t^{(2)}\right)\right\} \mid M_0, X_0\right] \\
&= P\left(M_t \in \left(\Phi_t^{(1)}\left(X_t^{(1)}\right), \Phi_t^{(2)}\left(X_t^{(2)}\right)\right] | M_0, X_0\right) > 0,
\end{aligned}$$

where to go from the first to the second line, we integrated the relationship (66) across all sample paths, and, in the last line, we used the fact that along any sample path, $\left(\Phi_t^{(1)}\left(X_t^{(M,1)}\right), \Phi_t^{(2)}\left(X_t^{(M,2)}\right)\right]$ has non-empty interior. This establishes the result for the IRF of the adoption decision, a_t. The result for the IRF of the user base follows from the relationship: $\mathcal{I}_X(t,S,X) = k\int_0^t \mathcal{I}_a(u,S,X)du + e^{-kt}X$. □

Proof of Predictions 1b and 2b. By Result 2, for any $C > 0$, the equilibrium sequence of adoption thresholds associated with C, which we denote by Φ, satisfies $\Phi \succ \Phi^{(0)}$, where $\Phi^{(0)} = \{\Phi_t^{(0)}\}_{t\geq 0} = \{\underline{M}_t\}_{t\geq 0}$ is the equilibrium sequence of adoption thresholds in the model where $C = 0$ (where $\underline{M}_t$ is defined in Lemma 3). Applying Lemma 13 to these two equilibrium sequences of thresholds establishes the results. □

Finally, we prove Prediction 3b.

Proof of Prediction 3b. For a given initial condition M_0, fix a particular sample path for cash-based demand, $M = \{M_t\}_{t\geq 0}$. Let $X_0^{(1)} < X_0^{(2)}$ be two initial sizes of the user base, and let the user base $\{X_t^{(M,i)}\}$ generated by the sample path M using the equilibrium adoption threshold Φ given initial condition $X_0^{(i)}$ be defined as:

$$X_t^{(M,i)} = k\int_0^t a_s(M_s, X_s^{(M,i)})ds + e^{-kt}X_0^{(i)}, \quad i = 1,2,$$

where $a_t(M,X) \equiv \mathbf{1}\{M \leq \Phi_t(X)\}$. We next show by induction that $X_t^{(M,1)} < X_t^{(M,2)}$ for all $t \geq 0$. Consider a particular date t and assume that $X_t^{(M,1)} < X_t^{(M,2)}$. We have:

$$X_{t+dt}^{(M,i)} = (1-kdt)X_t^{(M,i)} + a_t(M_t, X_t^{(M,i)})kdt, \quad i = a,b. \quad (67)$$

By Result 2, $\Phi_t(.)$ is increasing. Therefore, $\Phi_t\left(X_t^{(M,1)}\right) \leq \Phi_t\left(X_t^{(M,2)}\right)$, so that $a_t(M_t, X_t^{(M,2)}) \geq a_t(M_t, X_t^{(M,1)})$. Equation (67) then implies that $X_{t+dt}^{(M,1)} < X_{t+dt}^{(M,2)}$. Since $X_0^{(M,1)} < X_0^{(M,2)}$, we therefore have that $X_t^{(M,1)} < X_t^{(M,2)}$ for all $t \geq 0$. This proof also shows that along any sample path M,

$$a_t(M_t, X_t^{(M,1)}) \leq a_t(M_t, X_t^{(M,2)}),$$

which, for $M_0 = (1-S)M^c$, implies that $\mathcal{I}_a\left(t,S,X_0^{(1)},C\right) \leq \mathcal{I}_a\left(t,S,X_0^{(2)},C\right)$. Next we show that the inequality is strict when $C > 0$. In that case, by Result 2, $\Phi_t(.)$ is strictly increasing, so that for all $t \geq 0$,

$\Phi_t(X_t^{(1)}) < \Phi_t(X_t^{(2)})$. If:

$$\Phi_0(X_0^{(1)}) < M_0 \leq \Phi_0(X_0^{(2)}),$$

then:

$$\mathbb{E}_0\left[a_0(M_0, X_0^{(1)})|M_0, X_0 = X_0^{(1)}\right] = 0 < 1 = \mathbb{E}_0\left[a_0(M_0, X_0^{(2)})|M_0, X_0 = X_0^{(2)}\right],$$

which implies that the inequality is strict. Otherwise, the adoption decisions on impact are the same for the two initial values of the user base: $a_0(M_0, X_0^{(1)}) = a_0(M_0, X_0^{(2)}) \equiv \tilde{a}_0 \in \{0, 1\}$. In the infinitesimal time period $[0, dt]$, the user base is locally deterministic, and given by:

$$X_{dt}^{(1)} = (1 - kdt)X_0^{(1)} + \tilde{a}_0 kdt, \quad i = a, b, \tag{68}$$

so that $X_{dt}^{(1)} < X_{dt}^{(2)}$, and so $\Phi_{dt}\left(X_{dt}^{(1)}\right) < \Phi_{dt}\left(X_{dt}^{(2)}\right)$. Since $X_{dt}^{(i)}$, $i = 1, 2$, is known at time 0, we have:

$$\begin{aligned}
\mathbb{E}_0\left[a_{dt}(M_{dt}, X_{dt}^{(2)})\right] - \mathbb{E}_0\left[a_{dt}(M_{dt}, X_{dt}^{(1)})\right] &= P\left(M_{dt} \leq \Phi_{dt}\left(X_{dt}^{(2)}\right)\right) - P\left(M_{dt} \leq \Phi_{dt}\left(X_{dt}^{(1)}\right)\right) \\
&= P\left(\Phi_{dt}\left(X_{dt}^{(1)}\right) < M_{dt} \leq \Phi_{dt}\left(X_{dt}^{(2)}\right)\right) > 0,
\end{aligned}$$

establishing the result for $C > 0$. When $C = 0$, by Lemma 3, we know that the adoption threshold is $\Phi_t = \underline{M}_t$ which is independent of X_t. So the IRF of the adoption decision is given by $\mathcal{I}_a(t, S, X_0, 0) = P(M_t \leq \underline{M}_t)$. This expression is independent of X_0, establishing the result. □

B.5 Microfoundation with two-sided market

This appendix describes a version of the model with extended microfoundations. Relative to the baseline model, the model described here has two additional features. First, firms that have adopted the electronic payments technology can still accept payments in cash, so that the electronic payments technology is an add-on, not an alternative to cash. (That is, the model accommodates multihoming by firms.) Second, the choice of consumers between cash and electronic payments is explicitly modelled. (We also allow for multihoming by consumers.) The main result is that the model with extended microfoundations is isomorphic to the model described in Section 3.

Consumers There is a continuum of mass 1 of identical households. Each period, households randomly meet with firms. Each household holds D units of deposits, where D is exogenous and fixed. Deposits can be used for payment in retail transactions, either by converting them to cash or by using them in electronic payments. Households can only withdraw up to L_t units of cash, where L_t is exogenous. Finally, they behave myopically: each period, after observing the number of firms that accept electronic payments, $X_t \equiv \int_{i\in[0,1]} \mathbf{1}\{x_{i,t} = e\}\, di \in [0, 1]$, they solve the following problem:

$$\begin{aligned}
\max_{C_t^c, C_t^e, L_t^c, L_t^e} \quad & X_t\left(\zeta C_t^e + (1-\zeta)C_t^c\right) + (1 - X_t)C_t^c - \frac{1}{2\gamma}\left(\frac{L_t^e - L^e}{P_t}\right)^2 \\
\text{s.t.} \quad L_t^c + L_t^e &\leq D \quad [\lambda_t] \\
L_t^c &\leq L_t \quad [\mu_t] \\
P_t C_t^c &\leq L_t^c \quad [\nu_t^c] \\
P_t C_t^e &\leq L_t^e \quad [\nu_t^e]
\end{aligned}$$

Because meetings are random, the probability that a household meets a firm that accepts both electronic payments and cash is X_t. Upon meeting, the household and the firm decide on which means of payment to use in order to conduct the transaction. We assume that electronic money is chosen with probability ζ, and cash is chosen otherwise; the probability ζ is exogenous and constant. Meeting a firm that accepts both electronic payments and cash thus yields expected utility $\zeta C_t^e + (1-\zeta)C_t^c$ to the household. If the household instead meets a firm that only accepts cash, the meeting yields utility C_t^c.

Additionally, there are quadratic utility costs associated with holding real balances of electronic means of payment away from an exogenous level L^e. Here, L^e could be arbitrarily small. This cost is non-pecuniary: it is a shorthand for modeling cognitive or, in this static framework, opportunity costs of adjusting real balances of electronic money. Finally, the household's problem is subject to two constraints that state that consumption using either type of payment cannot exceed real balances of each type.[104] We assume that prices of consumption goods are constant, and normalize them to $P_t = 1$. Eliminating the multipliers ν_t^c and ν_t^e, the necessary first-order conditions for optimality for this problem can be written as:

$$\begin{aligned} &\lambda_t + \frac{1}{\gamma}\left(L_t^e - L^e\right) = \zeta X_t \\ &\lambda_t + \nu_t = 1 - X_t + (1-\zeta)X_t \end{aligned} \tag{69}$$

along with two complementary-slackness conditions, $\lambda_t\left(D - L_t^c - L_t^e\right) = 0$ and $\mu_t\left(L_t - L_t^c\right) = 0$. The two state variables of the household's problem are X_t and L_t.

Firms The problem of each firm is identical to that described in Section 3, except for the definition of flow profits of each firms. Namely, we now assume that profits are now given by:

$$\Pi(x_{i,t}, C_t^c, C_t^e) = \begin{cases} (\mu-1)\left(\zeta C_t^e + (1-\zeta)C_t^c\right) & \text{if } x_{i,t} = e, \\ (\mu-1)C_t^c & \text{if } x_{i,t} = c. \end{cases}$$

where $\mu > 1$ is a constant markup over marginal cost. Each period, the firm meets a different household. If the firm accepts electronic payments ($x_{i,t} = e$), its expected revenue is $\zeta C_t^e + (1-\zeta)C_t^c$. Otherwise, its revenue is C_t^c. The rest of the firms' problem is identical. Following the same steps as in the main text, net adoption benefits follow:

$$B_t = \mathbb{E}_t\left[\int_{s\geq 0} e^{-(r+k)ds}(\mu-1)\zeta\left(C_{t+s}^e - C_{t+s}^c\right)ds\right] \tag{70}$$

implying that the state variables relevant to the adoption decision are now (C_t^c, C_t^e). Following the same steps as in the baseline model, the law of motion for the user base is now:

$$dX_t = (a_t(C_t^c, C_t^e) - X_t)k dt, \tag{71}$$

where:

$$a_t(C_t^c, C_t^e) = \mathbf{1}\left\{B_t\left(C_t^c, C_t^e\right) \geq 0\right\}. \tag{72}$$

We define the best response correspondence, $\tilde{a}$, as in Equation (10) in the main text.

There is a unique exogenous stochastic process in the model, L_t, the dynamics of which we leave unspecified for now. We focus on equilibria where the consumption and payments decisions of households are

[104] Because of the static nature of the household's problem, these are not, strictly speaking, "cash in advance" constraints.

Markov in the two aggregate states (L_t, X_t). We define equilibria as follows.

Definition 7 (Equilibrium)**.** *Given a stochastic process for L_t, an equilibrium is (a) household choice rules $C_t^c, C_t^e, L_t^c, L_t^e$, and their associated Lagrange multipliers, all of which are functions $\mathbb{R}^2 \to \mathbb{R}$; (b) a set of adoption rules $a = \{a_t\}_{t \geq 0}$, where each $a_t : \mathbb{R}^2 \to \{0, 1\}$, (c) a stochastic process X_t for the user base, such that:*

1. *$\forall (t, L_t, X_t) \in \mathbb{R}^+ \times \mathbb{R} \times [0, 1]$, the adoption rule is a symmetric best reponse to itself:*

$$\hat{a}_t(C_t^c(L_t, X_t), C_t^e(L_t, X_t); a_{|t}) = a_t(C_t^c(L_t, X_t), C_t^e(L_t, X_t))$$

2. *$\forall (t, L_t, X_t) \in \mathbb{R}^+ \times \mathbb{R} \times [0, 1]$, $C_t^c(L_t, X_t), C_t^e(L_t, X_t), L_t^c(L_t, X_t), L_t^e(L_t, X_t)$ and their associated Lagrange multipliers satisfy the first-order conditions given in Equation (69);*

3. *the user base X_t follows the law of motion in Equation (71).*

Isomorphism to baseline model Next, we show that this model is isomorphic to the baseline model described in Section 3. Specifically, we assume that deposits, D, are large relative to both cash in circulation and to potential demand for electronic payments: $D \gg L_t + L^e + \gamma\zeta$. In this case, any equilibrium has the following features. First, $\lambda_t = 0$, since the deposit constraint is slack when deposits are sufficiently high. Second, when $X_t > 0$, the constraint $L_t^e = C_t^e$ binds, so that:

$$C_t^e = L^e + \gamma\zeta X_t.$$

Moreover, $\mu_t^c = \nu_t = 1 - X_t + (1 - \zeta)X_t > 0$, so that $C_t^c = L_t$. Additionally, when $X_t = 0$, the solution is $C_t^e = L^e$ and $C_t^c = L_t$. The flow benefits of adoption are then given by:

$$\Pi_t^e - \Pi_t^c = (\mu - 1)\zeta\,(C_t^e - C_t^c) = (\mu - 1)\zeta\,(L^e + \gamma\zeta X_t - L_t)$$

Thus, the microfounded model produces identical dynamics to the model in the main text so long as:

$$C = (\mu - 1)\gamma\zeta^2, \quad M^e = (\mu - 1)\zeta L^e, \quad M_t = (\mu - 1)\zeta L_t.$$

where C, M^e and M_t are the exogenous parameters and processes described in Section 3. In this version of the model, the reduced-form parameter governing externalities, $C = (\mu - 1)\gamma\zeta^2$, is large either when the slope of adjustment costs for electronic money, which is given by $1/\gamma$, is low (so that households adjust their holdings of electronic money rapidly in response to changes in X_t), or ζ is high, so that when a match between households using e-money and firms accepting it occurs, e-money is likely to be the medium of exchange chosen.

We make two final remarks about microfoundations. First, Results 1 and 2 do not depend on the specific parametric assumptions made in the baseline model of Section 3. They hold more generally, so long as the relative flow payoff between electronic money and cash ($\Delta\Pi$, defined in Lemma 1), satisfies Assumptions A0, A3 and A4 on Lipschitz continuity, strategic complementarities, and payoff monotonicity described Appendix B.1.[105] Therefore, a model with more general microfoundations (regarding, in particular, consumer utility

[105]In particular, linearity of payoffs is not required. Of course, the analytical expressions for strict dominance bounds derived in Lemma 3 may not hold.

or firm profits), so long as they satisfy these assumptions, would lead to the same qualitative predictions regarding endogenous persistence and state-dependence.

Second, it may not be immediately clear why, in the model with multihoming, firms may choose to give up the option to accept electronic payments, and return to cash. The reason is that, so long as $\zeta > 0$, firms expect, with strictly positive probability, to have to settle some transactions with electronic money. When cash-based demand is sufficiently high, compared to electronic payments (for instance, if $L^c \gg L^e$), doing so leads to an implicit opportunity cost of accepting electronic payments. This easiest to see when there are no complementarities, which corresponds to $\gamma = 0$ in the two-sided market model. In that case, households hold exactly $L^e_t = L^e$ balances of electronic money, so that $C^e_t = L^e$. The flow payoff from multihoming, relative to only accepting cash, is then $(\mu - 1)\zeta(L^e - L_t)$, which can be negative for sufficiently large values of L_t. With non-immediate adjustment, firms might therefore find it preferrable to move back to accepting only cash if L_t is sufficiently large.

B.6 Model with fixed cost

This section describes a model where electronic money has zero positive external returns, but its adoption requires that firms pay a fixed cost. We first describe the model and its solution. We then highlight how Predictions 1a-3a change in this model, compared to the model with external returns.

B.6.1 Model exposition

Description Each firm $i \in [0, 1]$ must choose between operating using one of two payment technologies, $\{e, c\}$, where e stands for electronic money, and c stands for cash. $x_{i,t} \in \{e, c\}$ is the technology choice of firm i at time t. For each firm, flow profits per unit of time are given by:

$$\Pi(x_{i,t}, M_t, X_t) = \begin{cases} M_t & \text{if } x_{i,t} = c, \\ M^e & \text{if } x_{i,t} = e, \end{cases} \tag{73}$$

where cash-based demand $\{M_t\}_{t \geq 0}$ follows:

$$dM_t = \theta \left(M^c - M_t\right) dt + \sigma dZ_t. \tag{74}$$

Note that, in order to be able to express the model solution in closed form, we have taken the limit $T \to +\infty$ of our baseline model, so that fundamentals are mean-reverting at rate θ regardless of the horizon. This in turn means that value and policy functions are stationary.

Firms discount the future at rate r. As in the baseline model, a firm may change the technology it uses to accept payments. This change is governed by a Poisson process with controlled intensity $\tilde{k}$ per unit of time — the "switching rate". In an infinitesimal period $(t, t + dt)$, a firm changes its payment technology with probability $\tilde{k}dt$, and keeps using the same technology with probability $(1 - \tilde{k}dt)$. The switching rate $\tilde{k}$ can be continuously adjusted by the firm, at no cost, subject to the constraint that $\tilde{k} \in [0, k]$, where k is an exogenous and fixed parameter, common to all firms.[106]

The key assumption regarding fixed costs is the following. If the firm is currently using electronic payments ($x_{i,t} = e$), and receives the Poisson shock to change its payment technology, it does not incur a

[106] We maintain the assumption that $k < +\infty$ so as to ensure comparability with the baseline model, but it is not required for the model to have a solution.

fixed cost. On the other hand, if it is currently using cash ($x_{i,t} = c$) and receives the shock, it must pay a fixed cost $\kappa > 0$. Thus, while the Bellman equation for the value of a firm with technology choice $x_t = e$ is the same as the one reported in the proof of Lemma 1 (except that it is now independent of X_t), the Bellman equation for the value of a firm with technology choice $x_t = c$ is:

$$\begin{aligned} V(c, M_t) &= \max_{\tilde{k}\in[0,k]} \Big\{ \Pi(c, M_t)dt \\ &\quad + \tilde{k}dt(1 - rdt)\mathbb{E}_t \left[V(e, M_{t+dt}) - \kappa\right] \\ &\quad + (1 - \tilde{k}dt)(1 - rdt)\mathbb{E}_t \left[V(c, M_{t+dt})\right] \Big\}, \end{aligned}$$

where the term $-\kappa$ in the second line reflects the payment of the fixed cost.

Equilibrium and aggregation An equilibrium of the model is simply defined as a set (stationary) optimal policies $\tilde{k}(x, M)$ and value functions $V(x, M)$ that satisfy the Bellman equations for firms with $x = e$ and $x = c$. We show below that optimal policies take the following generic form: there exist two boundaries $M_s \le M_S$ such that c-firms adopt e when $M_t < M_s$, e-firms adopt c when $M_t > M_S$, and are inactive when $M_t \in [M_s, M_S]$:

$$\tilde{k}(e, M_t) = \begin{cases} 0 & \text{if} \quad M_t \le M_S \\ k & \text{if} \quad M_t > M_S \end{cases} \tag{75}$$

$$\tilde{k}(c, M_t) = \begin{cases} 0 & \text{if} \quad M_t \ge M_s \\ k & \text{if} \quad M_t < M_s \end{cases} \tag{76}$$

Since we want to compare the size of the user base and the average adoption decisions across firms, we define:

$$a(e, M) = \mathbf{1}\{M_t \le M_S\}, \quad a(c, M) = \mathbf{1}\{M_t \le M_s\}. \tag{77}$$

The law of motion for X_t is then given by:

$$\begin{aligned} dX_t &= -X_t(1 - a(e, M_t))kdt + (1 - X_t)a(c, M_t)kdt \\ &= \Big\{a(c, M_t) - \Big(a(c, M_t) + (1 - a(e, M_t))\Big)X_t\Big\} kdt \end{aligned} \tag{78}$$

When $\kappa = 0$, $M_s = \underline{M}$, $M_S = \underline{M}$, $a(c, M_t) = a(e, M_t)$, and the model has the same law of motion as in the baseline model with no positive external returns, $C = 0$.[107]

B.6.2 Model solution

Following the same steps as in the proof of Lemma 1, we obtain:

$$\begin{aligned} rV(e, M_t)dt &= \Pi(e, M_t)dt + \mathbb{E}_t\left[dV(e, M_t)\right] - \max_{\tilde{k}\in[0,k]} \tilde{k}dtB(M_t) \\ rV(c, M_t)dt &= \Pi(c, M_t)dt + \mathbb{E}_t\left[dV(c, M_t)\right] + \max_{\tilde{k}\in[0,k]} \tilde{k}dt\left(B(M_t) - \kappa\right), \end{aligned}$$

[107] Here, $\underline{\mathrm{M}}$ is defined in Lemma 3, taking the limit $T \to +\infty$.

where:

$$B(M_t) \equiv V(e, M_t) - V(c, M_t).$$

The optimal arrival rates now depend on the current technology choice of the firm. They are given by:

$$\tilde{k}(e, M_t) = \begin{cases} 0 & \text{if} \quad B(M_t) \geq 0 \\ k & \text{if} \quad B(M_t) < 0 \end{cases} \tag{79}$$

$$\tilde{k}(c, M_t) = \begin{cases} 0 & \text{if} \quad B(M_t) \leq \kappa \\ k & \text{if} \quad B(M_t) > 0 \end{cases} \tag{80}$$

where we have assumed that if $B(M_t) = \kappa$, a firm that is currently using cash decides to stay with cash, and likewise, if $B(M_t) = 0$, a firm currently using electronic payments decides to stay with electronic payments. There is now an inaction region:

$$B(M_t) \in [0, \kappa] \implies \tilde{k}(e, M_t) + \tilde{k}(c, M_t) = 0.$$

In the region where $B(M_t) \in [0, \kappa]$, by taking the difference between the two Bellman equations characterizing the value of the firm, we see that the value of adoption satisfies the Bellman equation:

$$rB(M_t)dt = \Delta\Pi(M_t)dt + \mathbb{E}_t \left[dB(M_t) \right].$$

Taking the limit as $dt \to 0$, B must solve the ordinary differential equation:

$$\frac{1}{2}\sigma^2 B''(M) + \theta(M^c - M)B'(M) - rB(M) = M - M^e. \tag{81}$$

A particular solution to this equation is:

$$B_{P,b}(M) = \frac{1}{r + \theta} \left(M^c - M \right) - \frac{1}{r} \left(M^c - M^e \right), \tag{82}$$

and general solutions take the form:

$$B_b(M) = B_{1,b}\Phi\left(\frac{r}{2\theta}, \frac{1}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{2,b}\frac{\sqrt{\theta}}{\sigma}\left(M^c - M\right)\Phi\left(\frac{r+\theta}{2\theta}, \frac{3}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{P,b}(M), \tag{83}$$

where $\Phi(a, b; z)$ is Kummer's function. In the region where $B(M_t) < 0$, the differential equation becomes:

$$\frac{1}{2}\sigma^2 B''(M) + \theta(M^c - M)B'(M) - (r + k)B(M) = M - M^e. \tag{84}$$

A particular solution to this equation is:

$$B_{P,c}(M) = \frac{1}{r + k + \theta} \left(M^c - M \right) - \frac{1}{r + k} \left(M^c - M^e \right), \tag{85}$$

and general solutions take the form:

$$B_c(M) = B_{1,c}\Phi\left(\frac{r+k}{2\theta}, \frac{1}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{2,c}\frac{\sqrt{\theta}}{\sigma}\left(M^c - M\right)\Phi\left(\frac{r+k+\theta}{2\theta}, \frac{3}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{P,c}(M). \tag{86}$$

Finally, in the region where $B(M_t) > \kappa$, the differential equation becomes:

$$\frac{1}{2}\sigma^2 B''(M) + \theta(M^c - M)B'(M) - (r+k)B(M) = M - M^e - k\kappa. \tag{87}$$

A particular solution to this equation is:

$$B_{P,a}(M) = \frac{1}{r+k+\theta}\left(M^c - M\right) - \frac{1}{r+k}\left(M^c - \left(M^e + k\kappa\right)\right), \tag{88}$$

and general solutions take the form:

$$B_a(M) = B_{1,a}\Phi\left(\frac{r+k}{2\theta}, \frac{1}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{2,a}\frac{\sqrt{\theta}}{\sigma}\left(M^c - M\right)\Phi\left(\frac{r+k+\theta}{2\theta}, \frac{3}{2}; \frac{\theta}{\sigma^2}\left(M^c - M\right)^2\right) + B_{P,a}(M). \tag{89}$$

The value of adoption is then given by:

$$B(M) = \begin{cases} B_a(M) & \text{if} \quad M \leq M_s \\ B_b(M) & \text{if} \quad M \in [M_s, M_S] \\ B_c(M) & \text{if} \quad M \geq M_S \end{cases} \tag{90}$$

where the six coefficients $\{B_{1,a}, B_{1,b}, B_{1,c}, B_{2,a}, B_{2,b}, B_{2,c}\}$ and the two thresholds (M_s, M_S) satisfy the following eight conditions:

$$\begin{aligned}
&\lim_{M \to -\infty} B_a(M) = +\infty, \\
&B_a(M_s) = \kappa, \quad B_b(M_s) = \kappa, \\
&B'_a(M_s) = B'_b(M_s), \\
&B_b(M_S) = 0, \quad B_c(M_S) = 0, \\
&B'_b(M_S) = B'_c(M_S), \\
&\lim_{M \to +\infty} B_c(M) = -\infty.
\end{aligned} \tag{91}$$

B.6.3 Empirical predictions

We now discuss whether the three main empirical predictions developed in Section 3 for the model with positive external returns also apply to the model with fixed costs. We start with the predictions on endogenous persistence. We focus on the perfect forecast response of the economy to a shock at time 0, that is sufficiently large that:

$$M_0 = (1 - S)M_c < M_s.$$

As before, we define the perfect forecast response as the sample path for (M_t, X_t) which the innovations to M_t are exactly zero for all $t > 0$, so that cash demand follows:

$$\forall t \geq 0, \quad M_t = (1 - Se^{-\theta t})M_0.$$

The main predictions of the model regarding the persistence of the perfect foresight response are the following.

Prediction 1c. ***(Persistence in the response of the user base)*** *Assume that $M_S > M^c$. Following the shock, the user base increases permanently:*

$$\lim_{t \to +\infty} X_t > X_0. \tag{92}$$

Prediction 2c. ***(No persistence in the response of the adoption decision)*** *Following the shock, the adoption decision of firms currently using cash is given by:*

$$a(c, M_t) = \begin{cases} 1 & \text{if} \quad t \leq \hat{t}(S) \\ 0 & \text{if} \quad t > \hat{t}(S) \end{cases}, \tag{93}$$

where the horizon $\hat{t}(S)$ only depends on the persistence of cash demand, θ, and on the size of the shock relative to $M_0 - M_s$.

Appendix Figure H.15 illustrates these two predictions. In this figure, the adoption thresholds (M_s, M_S) are chosen to that $M_s < M^c < M_S$. The figure displays the perfect foresight trajectory of the economy following the shock, starting from a user base of $X_0 = 0$. The economy enters the adoption region at $t = 0^+$. So long as the economy is in that region, the trajectory of the user base is given by:

$$X_t = 1 - e^{-kt},$$

so that the user base increases as cash reverts towards its long-run mean. As a result, the economy moves up and to the right. At time:

$$\hat{t}(S) = \frac{1}{\theta} \log \left(\frac{SM^c}{M^c - M_s} \right),$$

the economy reaches the lower boundary of the inaction region. After this, given that $M_c < M_S$, the user base is given by:

$$\forall t \geq \hat{t}(S), \quad X_t = X_{\hat{t}(S)} = 1 - \left(\frac{M_0 - M^c}{SM^c} \right)^{\frac{\theta}{k}} > 0 = X_0.$$

Thus, in this example, the shock has a permanent effect on the user base. Note that the result of a permanent effect relies on the assumption that $M_S > M^c$. In turn, this generally requires that fixed adoption costs κ are sufficiently large.[108] In the case where $M_S < M^c$, the response of the user base need not be permanent, but, at any finite horizon, the user base will be larger than if fixed costs were zero, because the economy will always spend a strictly positive amount of time in the inaction region. Thus, as in the model with positive external returns, the user base responds to the shock more persistently than cash-based demand.

However, this persistence does not extend to the response of the adoption decision, contrary to the model with positive external returns. This is the point of Prediction 2c, which shows that firms that currently use

[108] Recall that as $\kappa \to 0$, the adoption thresholds converge to $M_S = M_s = \underline{M} < M^c$.

cash stop adopting electronic money at date $\hat{t}(S)$. Beyond that date, the shock has no effect on adoption decisions, and no firms currently using cash seek to adopt electronic money anymore.

We conclude by highlighting the fact that with fixed costs, there is no state-dependence in adoption decisions with respect to the initial user base.

Prediction 3c. ***(No state-dependence with respect to the initial user base)*** *The response of the adoption decision of firms currently using cash, $a(c, M_t)$, is independent of X_0.*

This is left without proof, since it immediately follows from the observation that all policy functions in the model are independent of the user base, X_t. Intuitively, in Appendix Figure H.15, because the adoption threshold does not depend on the user base, the dynamics of the economy following a large shock would be similar between regardless of the initial value of the user base, $X_0 > 0$.

B.7 Comparative statics with respect to volatility when $C = 0$

In this section, we some properties of the model without complementarities and highlight their implications for the model's comparativ statics with respect to uncertainty. Throughout, we take the limit $T \to +\infty$, so that there is no time-dependence in value and policy functions. This limit is well-defined in the case $C = 0$, since the strict dominance bounds of Lemma 3 coincide in this case. When $C = 0$, adoption decisions and the value of adoption are given by:

$$\begin{aligned} a(M_t) &= \mathbf{1}\{M_t < \underline{M}\}, \\ B(M_t) &= A_M(\underline{M} - M_t), \end{aligned} \tag{94}$$

where:

$$\underline{M} = M^c - \left(1 + \frac{\theta}{r+k}\right)(M^c - M^e),$$
$$A_M = \frac{1}{r+k+\theta}.$$

These expressions follow directly from Lemma 3. Adoption follows a threshold rule, where the threshold $\underline{M}$ is fixed and independent of X_t. The impulse response function (IRF) of the adoption decision, starting from $M_0 = M^c$, is given by:

$$\forall t \geq 0, \quad \mathcal{I}_a(t; X) \equiv \mathbb{E}\left[a_t \mid X_0 = X, M_0 = M^c\right]. \tag{95}$$

Note that contrary to the rest of the analysis, here we do not assume that there is a shock to the first moment of cash demand at time 0, so that cash demand $M^0 = M^c$ is at its long-run level. This helps focus the discussion on the effects of uncertainty, but does not change the two results highlighted below. Using Lemma 3, we then have that:

$$\mathcal{I}_a(t; X) = \begin{cases} 0 & \text{if} \quad t = 0 \\ F\left(\frac{-\nu^*}{\sigma_t} M^c\right) & \text{if} \quad t > 0 \end{cases} \tag{96}$$

where:

$$\nu^* \equiv \left(1 + \frac{\theta}{r+k}\right)\frac{M^c - M^e}{M^c}, \quad \sigma_t^2 \equiv (1 - e^{-2\theta t})\frac{\sigma^2}{2\theta}, \tag{97}$$

and $F(.)$ is the CDF of the standard normal distribution. This expression has the following two implications:

$$\begin{aligned}\forall t>0, \quad \frac{\partial}{\partial \sigma}\mathcal{I}_a(t;X) &= \frac{\sqrt{1-e^{-2\theta t}}}{\sqrt{2\theta}}\frac{\nu^* M^c}{\sigma_t^2} f\left(\frac{-\nu^* M^c}{\sigma_t}\right) > 0, \\ \frac{\partial^2}{\partial\sigma\partial X}\mathcal{I}_a(t;X) &= 0. \end{aligned} \tag{98}$$

The first expression implies that if uncertainty is larger and $\nu^* > 0$, which is equivalent to $M^c - M^e > 0$ (an assumption we maintain throughout the paper), then all else equal, higher uncertainty is associated with a higher probability of the economy being in the adoption region. This is because the likelihood of a large, negative shock to cash demand becomes higher. The second expression indicates that the effect of uncertainty (in a comparative statics sense) is independent of the initial size of the adoption base. In this particular sense, changes in the level of aggregate uncertainty should not be subject to the type of state-dependence we highlight in Predictions 2a and 3b.

Finally, the following Lemma helps characterize the autocorrelation of the adoption decision in the model without complementarities. This Lemma indicates that, as uncertainty increases, in the model without complementarities, the autocovariance of the adoption decision declines. This implies that responses to any given shock should be less persistent under higher uncertainty.

Lemma 14 (Persistence in the model without complementarities)**.** *For all $t, s \geq 0$, we have:*

$$\begin{aligned}\nu(s,t;X) &\equiv cov\left(\mathbf{1}\{M_s \leq \underline{M}\}, \mathbf{1}\{M_t \leq \underline{M}\} \mid M_0 = M^c, X_0 = X\right) \\ &= \mathcal{C}(\mathcal{I}_a(s,X), \mathcal{I}_a(t,X); \rho_{s,t}) - \mathcal{I}_a(s,X)\mathcal{I}_a(t,X), \end{aligned} \tag{99}$$

Here, $\rho_{s,t} \equiv e^{-\theta(t-s)}\frac{\sigma_s}{\sigma_t}$, and $\mathcal{C}(u,v;\rho)$ is the bivariate normal copula with parameter ρ. This autocovariance function declines with σ:

$$\frac{\partial \nu}{\partial \sigma}(s,t;X) < 0. \tag{100}$$

Proof of Lemma 14. Recall that the expression for the bivariate normal copula is:

$$\mathcal{C}(u,v;\rho) = \begin{cases} \max(u+v-1,0) & \text{if} \quad \rho = -1 \\ F_2\left(F^{-1}(u), F^{-1}(v); \rho\right) & \text{if} \quad -1 < \rho < 1 \\ \min(u,v) & \text{if} \quad \rho = 1 \end{cases}, \tag{101}$$

where $F_2(a,b;\rho)$ is the CDF of the bivariate standard normal with correlation coefficient ρ:

$$F_2(a,b;\rho) = \int_{x\leq a, y \leq b} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dxdy. \tag{102}$$

First, we note that:

$$P(M_s \leq \underline{M}, M_t \leq \underline{M} \mid M_0 = M^c) = C(\mathcal{I}_a(s,X), \mathcal{I}_a(t,X); \rho_{s,t}).$$

This expression is Sklar's theorem for the bivariate normal random vector (M_s, M_t). This expression implies the formula reported for the autocovariance function. To simplify notation, define:

$$h(t) \equiv \frac{-\nu^*}{\sigma_t} M^c.$$

Using the expression for the bivariate normal copula, and for the functions $\mathcal{I}_a(s, X)$, and taking derivatives with respect to σ, we get:

$$\begin{aligned}\frac{\partial \nu}{\partial \sigma}(s,t;X) &= f\left(g(s)\right)\frac{\partial h}{\partial \sigma}(s)\Big(P(M_t \leq \underline{M}|M_s = \underline{M}, M_0 = M^c) - 1\Big) \\ &+ f\left(g(t)\right)\frac{\partial h}{\partial \sigma}(t)\Big(P(M_s \leq \underline{M}|M_t = \underline{M}, M_0 = M^c) - 1\Big),\end{aligned}$$

where f is the pdf of the standard normal distribution. Since for any $0 \leq s \leq t$, h is increasing in σ, this proves the result. □

C Numerical solution method

This appendix describes the numerical procedure to solve for equilibrium policies, construct impulse response functions, and construct the ergodic distribution of the model of Section 3.

The numerical procedure first relies on discretizing the model to finite time intervals, Δt. We then proceed in two broad steps. First, we solve the model for $t > T$, which becomes stationary (since $\theta = 0$), using iterated deletion of strictly dominated strategies. This approach produces the unique equilibrium policy function for sufficiently small Δt.[109] Second, for $t \leq T$, we proceed by backward induction, using the solution for $t = T + \Delta t$ as our starting point. The unique equilibrium policy function can then be used to construct the impulse response functions and the ergodic distribution of the model using standard methods.

C.1 Fundamentals process

Recall that the fundamentals process follows:

$$\begin{aligned} dM_t &= \theta_t \left(M^c - M_t\right) dt + \sigma dZ_t, \\ \theta_t &= \begin{cases} \theta & \text{if} \quad t \leq T, \\ 0 & \text{if} \quad t > T. \end{cases}\end{aligned}$$

Let Δt denote an small time interval; then,

$$\mathbb{E}\left[M_{t+\Delta t}|M_t = M\right] = \begin{cases} M + \theta(M^c - M)\Delta t + o(\Delta t) & \text{if} \quad t \leq T - \Delta t, \\ M + o(\Delta t) & \text{if} \quad t > T - \Delta t; \end{cases}$$

$$\mathbb{E}\left[M^2_{t+\Delta t}|M_t = M\right] = \begin{cases} \sigma^2\Delta t + M^2 + 2M\theta\left(M^c - M\right)\Delta t + o(\Delta t) & \text{if} \quad t \leq T - \Delta t, \\ \sigma^2\Delta t + o(\Delta t) & \text{if} \quad t > T - \Delta t. \end{cases}$$

[109] In practice, convergence to the unique equilibrium policy function in the discretized model is never an issue.

We construct a corresponding discrete process that matches these first two moments following the methodology described in Miao (2013). We construct a grid $\mathcal{M} = (M_i)_{i=1}^{N_M}$, where $N_M = 2N + 1$, that has step size h_M and that is centered at $M = M^c$:

$$\forall N_M \geq i \geq 2, \quad M_i - M_{i-1} = h_M,$$
$$M_{N+1} = M^c.$$

Define the probabilities of up, down, and no move of the discretized process as:

$$\begin{aligned} u_{i,t} &\equiv \mathbb{E}\left[M_{t+dt} = M_i + h_M | M_t = M_i\right], \\ d_{i,t} &\equiv \mathbb{E}\left[M_{t+dt} = M_i - h_M | M_t = M_i\right], \\ 1 - u_{i,t} - d_{i,t} &\equiv \mathbb{E}\left[M_{t+dt} = M_i | M_t = M_i\right]. \end{aligned}$$

Given (θ, σ) and discretization parameters $(N, \Delta t)$, we set $(h_M, u_{i,t}, d_{i,t})$ as follows:

$$u_{i,t} = \begin{cases} \dfrac{\sigma^2 + h_M \theta (M^c - M_i)}{2h_M^2}\Delta t & \text{if} \quad t \leq T \\ \dfrac{1}{2} & \text{if} \quad t > T \quad \text{and} \quad 2 \leq i \leq N_M - 1 \\ 0 & \text{if} \quad t > T \quad \text{and} \quad i \in \{1, N_M\} \end{cases},$$

$$d_{i,t} = \begin{cases} \dfrac{\sigma^2 - h_M \theta (M^c - M_i)}{2h_M^2}\Delta t & \text{if} \quad t \leq T \\ \dfrac{1}{2} & \text{if} \quad t > T \quad \text{and} \quad 2 \leq i \leq N_M - 1 \\ 0 & \text{if} \quad t > T \quad \text{and} \quad i \in \{1, N_M\} \end{cases},$$

$$h_M = \frac{\sigma}{\sqrt{N\theta}}.$$

These restrictions guarantee that $d_{1,t} = 0$, $u_{N_M,t} = 0$, and that the discretized process matches the two first moments of the continuous time process up to order $o(\Delta t)$ for $t \leq T$.

We make two additional restrictions. First, we require that $h_M = \sigma\sqrt{\Delta t}$, so that the discretized process also matches the two first moments up to order $o(\Delta t)$ for $t > T$. This requires:

$$\Delta t = (N\theta)^{-1}. \tag{103}$$

With this first additional restriction, the upper and lower bounds of $\mathcal{M}$ are:

$$M_1 = M^c - \sqrt{\frac{N}{\theta}}\sigma, \quad M_{N_M} = M^c + \sqrt{\frac{N}{\theta}}\sigma.$$

The second additional restriction is that the grid always contains the upper and lower bounds for the strict dominance regions, which are $\underline{M} = -\overline{w}$ and $\overline{M} = -\underline{w}$, where $\underline{w}$ and $\overline{w}$ are given in Appendix Table H.19. This requires that:

$$N \geq \sigma^2 \theta \max\left((\overline{M} - M^c)^2, (M^c - \underline{M})^2\right). \tag{104}$$

Note that with these restrictions, for all $i = 1, ..., N_M$, the expressions for the probabilities of up and down jumps for $t \leq T$ simplify to

$$u_{i,t} = \frac{1}{2}\left(1 + \frac{\sqrt{\theta}}{\sigma\sqrt{N}}\left(M^c - M_i\right)\right),$$

$$d_{i,t} = \frac{1}{2}\left(1 - \frac{\sqrt{\theta}}{\sigma\sqrt{N}}\left(M^c - M_i\right)\right).$$

As $N \to +\infty$, $\Delta t \to 0$, and the discretized process converges to the continuous-time process.

C.2 Discrete approximation

The discrete-time counterpart to Equation (7), the continuous-time Bellman equation defining the relative value of adoption, is:

$$B_t(M_t, X_t) = \Pi(M_t, X_t)\Delta t + (1 - (r + k)\Delta t)\mathbb{E}_t\left[B_{t+\Delta t}(M_{t+\Delta t}, X_{t+\Delta t})|M_t, X_t\right], \tag{105}$$

where the law of motion for X_t follows:

$$X_{t+\Delta t} = X_t + a_t(M_t, X_t)(1 - X_t)k\Delta t - (1 - a_t(M_t, X_t))X_t k\Delta t, \tag{106}$$

where $a_t(M_t, X_t)$ is the adoption rule. Note that in this discrete approximation, we assume that X_t is locally deterministic (that is, $X_{t+\Delta t}$ is known at time t), since we only allow a_t to depend on (M_t, X_t), consistent with the continuous-time model.

We then solve for the sequence of functions $\{B_t\}_{t\geq 0}$ and $\{a_t\}_{t\geq 0}$ in this discrete approximation in two steps. First, for $t > T$, $\theta_t = 0$, so the model becomes stationary and a and B do not depend on time. Additionally, the equilibrium is unique. In order to find this equilibrium, we proceed by upper and lower deletion of strictly dominated strategies, in a manner which we describe below. Once this solution has been obtained, for $t \leq T$, we then proceed by backward induction, starting from the $a_{T+\Delta t} = a$ and $B_{T+\Delta t} = B$.

We solve for $\{B_t(.,.)\}_{t\geq 0}$ on $\mathcal{M}$ and, for X_t, on a grid $\mathcal{X} = (X_j)_{j=1}^{N_X}$ on $[0, 1]$, with step size $h_X = 1/N_X$. With some abuse of notation, $\{B_t\}_{t\geq 0}$ and $\{a_t\}_{t\geq 0}$ will refer to the $N_M \times N_X$ matrices characterizing the value of adoption and optimal adoption choices on these grids.

Preliminaries Define the sequence of $N_M \times N_M$ matrices $\left(J^{(M,t)}\right)_{t\geq 0}$ by:

$$J_{i,j}^{(M,t)} = \begin{cases} 1 - u_{i,t} - d_{i,t} & \text{if} \quad j = i, \\ u_{i,t} & \text{if} \quad j = i + 1 \text{ and } i \leq N_M - 1, \\ d_{i,t} & \text{if} \quad j = i - 1 \text{ and } i \geq 2. \end{cases} \tag{107}$$

Then, for any function $F : M \to F(M)$ defined on the grid $\mathcal{M}$ and represented by a vector F of size $N_M \times 1$, we have:

$$\mathbb{E}_t\left[F(M_{t+\Delta t})\right] = J^{(M,t)}F,$$

where $\mathbb{E}_t\left[F(M_{t+\Delta t})\right] \equiv \left(\mathbb{E}\left[F(M_{t+\Delta t})|M_t = M_i\right]\right)_{i=1}^{N_M}$.

Next, define the vectors:

$$dX^{(-)} \equiv \Big(-k\Delta t X_j\Big)_{j=1}^{N_X}, \quad dX^{(+)} \equiv \Big(k\Delta t(1-X_j)\Big)_{j=1}^{N_X}.$$

and the $N_X \times N_X$ matrices $J^{(X,+)}$ and $J^{(X,-)}$ by:

$$J_{i,j}^{(X,-)} = \begin{cases} 1 + \dfrac{dX_i^{(-)}}{h_X} & \text{if} \quad j = i, \\ -\dfrac{dX_i^{(-)}}{h_X} & \text{if} \quad j = i+1 \text{ and } i \leq N_M - 1, \end{cases}$$
$$J_{i,j}^{(X,+)} = \begin{cases} 1 - \dfrac{dX_i^{(+)}}{h_X} & \text{if} \quad j = i, \\ \dfrac{dX_i^{(+)}}{h_X} & \text{if} \quad j = i-1 \text{ and } i \geq 2. \end{cases} \tag{108}$$

For any function $F : X \to F(X)$ with values on the grid $\mathcal{X}$ given by a vector F of size $1 \times N_X$, the matrices $J^{(X,-)}$ and $J^{(X,+)}$ can be used to construct the linearly interpolated values of F at $X_{t+\Delta t}$ when $X_{t+\Delta t} = X_t - X_t k\Delta t$ and when $X_{t+\Delta t} = X_t + (1-X_t)k\Delta t$, respectively, by:

$$F^{(+)}(X_{t+\Delta t}) = FJ^{(X,-)} \quad \text{and } F(X_{t+\Delta t}) = FJ^{(X,+)},$$

where $F^{(+)}(X_{t+\Delta t})$ and $F^{(-)}(X_{t+\Delta t})$ both are vectors of size $1 \times N_X$.

Using these matrices, for any function $F : (M, X) \to F(M, X)$ with values on the grid $\mathcal{M} \times \mathcal{X}$ given by a matrix F of size $N_M \times N_X$, we can construct the two following approximate conditional expectations:

$$\mathbb{E}_t^{(-)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right] = J^{(M,t)} F J^{(X,-)},$$
$$\mathbb{E}_t^{(+)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right] = J^{(M,t)} F J^{(X,+)}.$$

Both are matrices of size $N_M \times N_X$. The (i, j) entry in $\mathbb{E}_t^{(-)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right]$ is the approximate expectation of $F(M_{t+\Delta t}, X_{t+\Delta t})$ conditional on $M_t = M_i$, $X_t = X_j$, and assuming that $X_{t+\Delta t} = X_t - X_t k\Delta t$. The expectation is approximate because F is linearly interpolated with respect to X. The entries of the matrix $\mathbb{E}_t^{(+)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right]$ have similar interpretations.

In order to construct the equilibrium, the impulse response functions, and the stationary distribution of the model, we will use the conditional expectations operator $\Gamma(.)$, which is a self-map on matrices of size $N_M \times N_X$, by:

$$\Gamma(F; A^{(+)}, J^{(X,+)}, A^{(-)}, J^{(X,-)}, J^{(M)}) = A^{(+)} \odot (J^{(M)} F J^{(X,+)}) + A^{(-)} \odot (J^{(M)} F J^{(X,-)}),$$

where $\odot$ is the Hadamard product. Here, the matrices $A^{(+)}$ and $A^{(-)}$ have entries in $\{0, 1\}$ and satisfy $A_{i,j}^{(+)} = 1 - A_{i,j}^{(-)}$ for all i, j. They encode a particular adoption rule: $A_{i,j}^{(t,+)} = a_t(M_i, X_j)$. Because the movements in X are locally deterministic (that is, $X_{t+\Delta t}$ depends on M_t, X_t), for any function $F : (M, X) \to F(M, X)$ with values on the grid $\mathcal{M} \times \mathcal{X}$ given by a matrix F of size $N_M \times N_X$, we then have:

$$\mathbb{E}_t\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right]$$

$$= a_t \odot \mathbb{E}_t^{(+)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right] + (1-a_t)\odot \mathbb{E}_t^{(-)}\left[F(M_{t+\Delta t}, X_{t+\Delta t})\right]$$

$$= \Gamma(F; A^{(t,+)}, J^{(X,+)}, A^{(t,-)}, J^{(X,-)}, J^{(M,t)})$$

Note that the equality only holds up to order $o(h_X)$ because the function F is interpolated with respect to its second argument when applying the operator Γ to compute the conditional expectation. Finally, we define the $N_M \times N_X$ matrix of incremental flow profits from e-money adoption Π as:

$$\Pi_{i,j} = M^e + CX_j - M_i \quad \forall 1 \le i \le N_M, \quad 1 \le j \le N_X.$$

Solution for $t > T$: upper and lower iterated deletion For $t > T$, the model is stationary. For easy of notation, we therefore omit time subscripts. Additionally, as discussed above, the solution to the continuous-time model is unique. We compute an approximation to this solution by applying iterated deletion of strictly dominated strategies to the discrete-time model.

Each iteration proceeds as follows. Let $(a^{(n)}, B^{(n)})$ be the $N_M \times N_X$ matrices representing the adoption strategy profile and value of adoption obtained after n iterations. Then, $(a^{(n+1)}, B^{(n+1)})$ are constructed as follows:

1. Define the matrices $A^{(n+1,+)}$ and $A^{(n+1,-)}$ as:

$$A_{i,j}^{(n+1,+)} = a_{i,j}^{(n)}, \quad A_{i,j}^{(n+1,-)} = 1 - A_{i,j}^{(n+1,+)}, \quad \forall 1 \le i \le N_M, \quad 1 \le j \le N_X$$

2. Compute:

$$B^{(n+1)} = \Pi\Delta t + (1-(r+k)\Delta t)\Gamma\left(B^{(n)}; A^{(n+1,+)}, J^{(X,+)}, A^{(n+1,-)}, J^{(X,-)}, J^{(M)}\right).$$

3. Compute:

$$a_{i,j}^{(n+1)} = \mathbf{1}\left\{B_{i,j}^{(n+1)} \ge 0\right\} \quad \forall 1 \le i \le N_M, \quad 1 \le j \le N_X.$$

We stop the iteration when $\max_{i,j} |B_{i,j}^{(n+1)} - B_{i,j}^{(n)}| < \delta$, when δ the convergence criterion. Note that the adoption value at iteration $(n+1)$ is computed assuming that other firms use the strategy profile $a^{(n+1)}$ resulting from the value of adoption obtained at iteration (n).

For the upper deletion of strictly dominated strategies, we start from the adoption value and strategy profiles:

$$B_{i,j}^{(0)} = A_M\left(\overline{\Phi}(X_j) - M_i\right), \quad a_{i,j}^{(0)} = 1, \quad \forall 1 \le i \le N_M, \quad 1 \le j \le N_X,$$

where the expressions for A_M and $\overline{\Phi}$ are given by Lemma 3 (using $\theta = 0$ and $T = +\infty$). This gives the adoption value implied by assuming that all firms adopt e-money. From this starting point, the sequences $\left(B_{i,j}^{(n)}\right)_{n\ge 0}$ and $\left(a_{i,j}^{(n)}\right)_{n\ge 0}$ are weakly decreasing. Since (by Lemma 3), they are bounded, they converge to a unique limit. For the lower deletion of strictly dominated strategies, we start from the adoption value and strategy profiles:

$$B_{i,j}^{(0)} = A_M\left(\underline{\Phi}(X_j) - M_i\right), \quad a_{i,j}^{(0)} = 0, \quad \forall 1 \le i \le N_M, \quad 1 \le j \le N_X,$$

which likewise generates a monotonically increasing sequence of adoption values and adoption rules that converge to a unique limit.

In the limit $\Delta t \to 0$, the two limits must coincide, since the equilibrium is known to be unique. Since the model only provides a discrete-time approximation (for which unicity is not guaranteed), we check numerically that the maximum across all states on $\mathcal{M} \times \mathcal{X}$ between the upper and lower limits of the iterated deletion sequences is less than δ, where δ is some convergence criterion.

Solution for $t \leq T$: backward induction Denote the solution obtained for $t \geq T + \Delta t$ as (a, B). For $t \leq T$, the solution of the continuous-time model is non-stationary, so we introduce time indexes again.

In order to compute the discrete-time approximation to the solution by backward induction, we make the following assumption:

$$a_t(M_t, X_t) = \mathbf{1}\left\{B_{t+\Delta t}\left(M_t, X_t\right) \geq 0\right\}.$$

In other words, we use the $t + dt$ optimal adoption rules to compute the time-t conditional expectation of adoption value. In the continuous-time limit, $\Delta t \to 0$, so the computed decision rule coincides with the equilibrium one, $a_t = \mathbf{1}\left\{B_t\left(M_t, X_t\right) \geq 0\right\}$. Additionally, as discussed below, we focus on horizons T that are sufficiently large such that a and B are time-invariant for small t, so that the computed decision rule $a_t(M_t, X_t) = \mathbf{1}\left\{B_{t+\Delta t}\left(M_t, X_t\right) \geq 0\right\}$ and the equilibrium decision rule $a_t(M_t, X_t) = \mathbf{1}\left\{B_t\left(M_t, X_t\right) \geq 0\right\}$ are identical up to numerical tolerance.

Backward induction proceeds as follows: given $(a_{t+\Delta t}, B_{t+\Delta t})$, we compute:

$$\begin{aligned}
A_t^{(+)} &= a_{t+dt}, \\
A_t^{(-)} &= 1 - A_t^{(+)}, \\
B_t &= \Pi \Delta t, \\
&+ (1 - (r+k)\Delta t)\Gamma\left(B_{t+\Delta t}; A^{(t,+)}, J^{(X,+)}, A^{(t,-)}, J^{(X,-)}, J^{(M,t)}\right), \\
a_t &= \mathbf{1}\left\{B_t \geq 0\right\}.
\end{aligned}$$

We initiate at date $t = T$ using the stationary value functions $(a_{T+\Delta t} = a, B_{T+\Delta t} = B)$. Finally, we check that the horizon of mean-reversion, T, is sufficiently large so that for all dates $0 \leq t \leq t_{max} - \Delta t$,

$$\max_{0 \leq t \leq t_{max} - \Delta t} \max_{i,j} \left|B_{t,i,j} - B_{t+\Delta t,i,j}\right| < \delta,$$

where δ is a converge criterion, and t_{max} a maximum horizon of analysis for the model. The guarantees that the backward induction has been repeated a sufficient number of periods for the solution to be stationary up to numerical tolerance on the time interval $0 \leq t \leq t_{max}$, so that the two adoption rules $a_t(M_t, X_t) = \mathbf{1}\left\{B_t\left(M_t, X_t\right) \geq 0\right\}$ are identical up to numerical tolerance for $0 \leq t \leq t_{max}$.

C.3 Impulse response functions

In the continuous-time model, the impulse response functions (IRFs) at horizon t are defined as:

$$\forall t \geq 0, \quad \mathcal{I}_a(t, M, X) \equiv \mathbb{E}\left[a_t \mid M_0 = M, X_0 = X\right],$$

$$\mathcal{I}_a(t, M, X) \equiv \mathbb{E}\left[X_t \mid M_0 = M, X_0 = X\right].$$

We next discuss the computation of the conditional expectation

$$\mathbb{E}\left[F_t(M_t, X_t) \mid M_0 = M, X_0 = X\right],$$

in the dsicrete approximation to the model, for any $t = n\Delta t$, $n \in \mathbb{N}$, and any deterministic sequence of functions with values on the grid $\mathcal{M} \times \mathcal{X}$ given by matrices $\{F_t\}_{t\geq 0}$ of size $N_M \times N_X$. For any $n \in \mathbb{N}$ and $t = n\Delta t$, we have:

$$\mathbb{E}_{t-\Delta t}\left[F_t(M_t, X_t)\right] = \Gamma(F_t; A^{(t-\Delta t,+)}, J^{(X,+)}, A^{(t-\Delta t,-)}, J^{(X,-)}, J^{(M,t-\Delta t)}),$$

where $\mathbb{E}_{t-\Delta t}\left[F(M_t, X_t)\right]$ is an $N_M \times N_X$ matrix. By applying the law of iterated expectations, we obtain:

$$\mathbb{E}_0\left[F_t(M_t, X_t)\right] = \Gamma_0 \circ \Gamma_{\Delta t} \circ ... \circ \Gamma_{t-\Delta t}(F_t),$$

where we used the shorthand Γ_t for the conditional expectations operator at time t:

$$\Gamma_t(F_{t+\Delta t}) \equiv \Gamma(F_{t+\Delta t}; A^{(t,+)}, J^{(X,+)}, A^{(t,-)}, J^{(X,-)}, J^{(M,t)}).$$

In our analysis of the model, we focus on IRF at horizons $t \leq t_{max}$, where t_{max} satisfies the following two conditions are satisfied:

$$t_{max} < T \quad \text{ and } \max_{0\leq t\leq t_{max}-\Delta t} \max_{i,j} |B_{t,i,j} - B_{t+\Delta t,i,j}| < \delta,$$

where δ is a convergence criterion. Under these conditions, up to numerical tolerance, the conditional expectations operator is constant over time: $\Gamma_t = \Gamma$. If the sequence of functions $\{F_t\}_{t_{max}\geq t\geq 0}$ is also constant over time, $F_t = F$, we have the relationship:

$$\mathbb{E}_0\left[F(M_t, X_t)\right] = \Gamma(\mathbb{E}_0\left[F(M_{t-\Delta t}, X_{t-\Delta t})\right]),$$

which we use to compute the IRF recursively. Using this relationship, for $\mathcal{I}_a$, we compute recursively:

$$\forall t_{\max}/\Delta t \geq n \geq 1, t = n\Delta t, \quad \mathcal{I}_a(n\Delta t) = \Gamma(\mathcal{I}_a((n-1)\Delta t)),$$

with $\mathcal{I}_a(0) = a_0$. Likewise, for $\mathcal{I}_X$, we compute

$$\forall t_{\max}/\Delta t \geq n \geq 1, t = n\Delta t, \quad \mathcal{I}_X(n\Delta t) = \Gamma(\mathcal{I}_X((n-1)\Delta t)),$$

with $\mathcal{I}_X(0) = M_X$, where M_X is an $N_M \times N_X$ matrix with all rows equal to $\mathcal{X}$.

C.4 Ergodic distribution

Finally, we use the conditional expectations operator to compute the stationary distribution of the model. We first express the conditional expectations operator in matrix form. Define:

$$\tilde{\Gamma}_t \equiv diag\left(vec\left(A^{(t,+)}\right)\right)\left(tr\left(J^{(X,+)}\right) \otimes J^{(M,t)}\right) + diag\left(vec\left(A^{(t,-)}\right)\right)\left(tr\left(J^{(X,-)}\right) \otimes J^{(M,t)}\right),$$

where tr is the transpose operator, vec is the vectorization operator, $diag(X)$ is the diagonal matrix with diagonal elements equal to the vector X, and $\odot$ is the Kronecker product. The matrix $\tilde{\Gamma}_t$ is a squared matrix of size $(N_M N_X) \times (N_M N_X)$. For any matrix B of size $N_M \times N_X$, we have that:

$$vec(\Gamma_t(B)) = \tilde{\Gamma}_t vec(B).$$

$\tilde{\Gamma}_t$ is a Markov transition matrix, that is, its rows sum up to 1. We now focus on dates $t \leq t_{max}$, where the operator is constant up to numerical tolerance: $\Gamma_t = \Gamma$, and $\tilde{\Gamma}_t = \tilde{\Gamma}$. We then diagonalize the transpose operator $tr(\tilde{\Gamma})$:

$$tr(\tilde{\Gamma})U = U diag(\lambda_1, ..., \lambda_{N_M N_X}),$$

where we normalize the columns of U so that $\sum_i |U_{i,n}| = 1$ for all n. The matrix $tr(\tilde{\Gamma})$ is the discrete equivalent of the Kolmogorov forward operator governing the law of motion of the distribution of the model over the states (M_t, X_t). Because $\tilde{\Gamma}$ is a Markov transition matrix, the first eigenvalue of $tr(\tilde{\Gamma})$ is $\lambda_1 = 1$. Define $\mu = U_1'$, where U_1 is the first column of U. Then $\mu tr(\tilde{\Gamma}) = \mu$. If at time $t = 0$, the distribution of the model is $\mu_0 = \mu$, we then have, for any $0 \leq t \leq t_{max}$, $t = n\Delta t$, $\mu_{n\Delta t} = tr(\tilde{\Gamma})^n \mu = \mu$, so that $\mu_t = \mu$ for all dates. Therefore, the distribution is constant up to numerical tolerance for $0 \leq t \leq t_{max}$.

D Cash contraction and Consumption

In this Section, we examine how household consumption responded to the cash swap using the same identification strategy from Section 4. In other words, we compare behaviors across districts that were characterized by different exposure to chest banks before the Demonetization. The objective of this analysis is twofold. First, these tests can provide novel evidence on how the Demonetization affected the real economy. Results from previous sections provide evidence that the Indian Demonetization led to a widespread and persistent rise in electronic payments. Given the size and speed of these responses, a natural question is whether the rise in electronic money was indeed sufficient to shield the real economy from the cash crunch.

Second, as discussed briefly in Section 4, this evidence on consumption is useful because it provides further robustness on the quality of our empirical model to identify the supply side effect of a cash contraction. The intuition for this second aspect is simple. Our tests on electronic payment — in particular the sharp response right around the policy shock — provides very strong evidence regarding the fact that our estimates capture how electronic payment use was affected by the Demonetization. However, the Demonetization could have affected the use of electronic payments in several ways, and not only because of a contraction in cash (supply shock). For instance, the Demonetization may have increased the overall uncertainty in the economy, which in turn may have reduced consumption.

The good news is that consumption response may help separating explanations based on cash contraction from alternative demand-side mechanisms. In particular, a demand side explanation would generally predict that the effects for consumption and electronic payments should go in the same direction. Instead, the opposite results — i.e. highly exposed areas experienced both higher increase in electronic payments and lower consumption — would be hard to rationalize by a demand mechanism, but easy to interpret as a supply side shock. In this sense, exploring the consumption response could provide useful evidence for our mechanism. In terms of robustness, the consumption data has a longer time series than the electronic payments. This will allow us to run several extra tests on the quality of our analysis.

D.1 Empirical setting

To measure the changes in consumption behavior by Indian households, we use data from the Consumer Pyramids database maintained by the Center for Monitoring Indian Economy (CMIE). This dataset has two crucial advantages relative to the widely used National Sample Survey (NSS), which is a consumption survey conducted by the central government agencies. First, the NSS is not available for the period of interest, as it was ran for the last time in 2011. Second, the NSS is a repeated cross-section of households, while CMIE data is a panel.

The data set provides a representative sample of Indian households, where households are selected to be representative of the population across 371 "homogeneous regions" across India. The survey has information on the monetary amount of the household expenditure across different large categories and some other background information on the members of the households. The expense categories include food, intoxicants, clothing and footwear, cosmetics and toiletries, restaurants, recreation, transport, power and fuel, communication and information services, health, education, bills and rents, appliances, equal monthly installments (EMIs), and others. Overall, the data quality is considered high, in particular since CMIE collects the data in person using specialized workers. Each household is interviewed every four months and is asked about their consumption pattern in the preceding four months. About 39,500 households are surveyed every month.

The data is organized in event-time around the month of the shock. In other words, for each household we aggregate data at the wave-level and we define the time of each wave relative to the wave containing November 2016. The final sample used in the analysis is constituted by about 95,000 households. We reach this count because we consider households for which the age of the head of household is between 18 and 75 years as of September 2016. To make the panel balance, we also only consider households with non-missing information between June 2016 and March 2017.

The main difference compared with the analyses in Section 4 is the timing. Before, the district-level data were measured at monthly level. For these household data, the survey procedure is such that households belonging to different waves of interviews are asked about the same month at different points in time. Therefore, the reporting on November 2016 — the first month of the shock — is generally clustered together with a different group of months depending on the wave.[110] This feature is quite common among consumer surveys, and it is similar to the Consumer Expenditure Survey in the US.[111] Following the literature in this area (e.g. Parker et al. (2013)), we deal with this feature by organizing the data by event-time. In other words, for each household we aggregate data at the wave-level and we define the time of each wave relative to the wave containing November 2016.[112]

With this data set of about 95,000 households, we then estimate the following household-level difference-in-difference model:

$$\log\left(y_{h,d,t}\right) = \alpha_t + \alpha_h + \delta_t\left(\text{Exposure}_d \times \mathbf{1}_{\{t \geq t_0\}}\right) + \Gamma_t' Y_{h,d} + \epsilon_{h,d,t}, \tag{109}$$

where $y_{h,d,t}$ are consumption measures for household h in district d and survey-time t, α_t and α_h are event-time and household fixed effects, Exposure_d is the district's exposure as described in Section 4, which is

[110] For example, 25% percent of households will be asked about August-November 2016 consumption in December 2016, 25% percent will be asked about September-December 2016 consumption in January 2017 and so on. Thus, November 2016 consumption will be recorded with other months depending on the month it was surveyed between December 2016-March 2017.

[111] The main difference is that the Consumer Expenditure Survey is run every three months rather than four months.

[112] Therefore, the time in the panel is the one for the wave in which a household was interviewed about November, and it is zero for the wave that happened four months before the one that includes November 2016 and one for the one that happened four months after.

interacted with dummies for the survey-time post-Demonetization, and $Y_{h,d}$ are controls, which are either at district or individual level. For controls in the regression, we use the same district-level covariates as in the previous set of analyses along with the addition of household-level controls including the age of the head of the household and log of household income, both measured as in the last survey before the shock. As usual, standard errors are clustered at district level, which is the level of the treatment.

D.2 Main results

Table H.20 shows the results for consumption responses based on exposure to the shock. Column (1) shows that relative to the pre-period, total consumption was cut more for households located in the highly affected district. The effect is sizable: a one-standard deviation increase in the chest bank score corresponds to about a 3.6% relative decline in total consumption. The same holds when using a dichotomous version of the shock: in this case, the highly affected households (top quartile) saw a relative drop of about 5.7%. Importantly, these results are not driven by differences in pre-trends between affected districts (Figure H.18).[113]

Therefore, the cash contraction negatively affected household consumption. However, there are three important things to point out about this negative effect. First, the impact of the shock was temporary. Looking at the interaction between the treatment and dummies identifying the next 3 waves in which the household was interviewed, we consistently find a small and non-significant coefficient. This effect suggests that the cash contraction only significantly impacted household behavior during the months immediately after the Demonetization and did not lead to a permanent change in consumption behavior. This evidence is consistent with the idea that the shock was really only binding between November and January.

Second, consistent with the idea that households were able to partially limit the impact of the shock, the contraction in consumption was larger for items that are less costly to cut for households. As a first step, we divide consumption into necessary and unnecessary items, where the former group contains expenses for food, rent and bills, and utilities (power and gas) while the latter contains the remaining part of the consumption basket. Table H.21 shows that, when consumption is split between the two baskets of goods, the effect on unnecessary consumption was economically larger (about 22% higher).[114]

This last result does not depend on the way we categorize consumption as necessary and unnecessary. In Columns (3)-(5) of Table H.21, we consider three consumption categories: rent and bills, food, and recreational expenses. For the first group - rent and bills - we find essentially no effect of the Demonetization. For food, the effect is still negative and significant. In particular, a one standard-deviation increase in exposure led to about 3% decline in food expenditure. However, this effect on food dwarfs in comparison to the cut on recreational expenses. For this category, we find that a standard-deviation increases led to more than a 15% cut in consumption.

Third, we also find direct evidence that electronic payments helped to partially limit the impact of the shock. While this evidence confirms that the rise in electronic payment was unable to undo the effects of the cash contraction, it may still be the case that electronic payments helped to partially limit the impact of the shock. To test this hypothesis, we examine the responsiveness to the shock across areas characterized by different levels of penetration of electronic payments in the pre-shock period. In particular, we focus on

[113] This analysis shows a positive and borderline significant effect on consumption two quarters after the Demonetization. One interpretation is that households have shifted some consumption to the future. Consistent with this interpretation, we actually find that the effect is driven entirely by unnecessary consumption, which is a category that contains durable expenditure. However, we also want to point out that this positive result is statistically weak and it does not replicate using alternative treatment specifications (e.g. using top quartile).

[114] The same difference also holds when looking using a dichotomous treatment (Appendix Table H.22): here necessary consumption is cut by 4%, while unnecessary consumption by about 8%.

the penetration of debit cards, which we proxy by the number of ATMs per million people in a district.[115] Our focus on traditional electronic payment is motivated by its relative size. In fact, debit cards represent the largest share of electronic transaction in India. Furthermore, while the issuance of new debit cards was overall modest, the Demonetization led to an increase in the amount of transactions, suggesting that debit cards were indeed used as a way to replace cash during the shock period.

The results of this analysis are presented in Table H.23. The key parameter in these regressions is the triple interaction between the time dummies, the measure of exposure to the shock, and a dummy that a value of one for districts that have an above-median number of ATMs per one million people. We repeat the same analysis using both the continuous (odd columns) and dichotomous (even columns) versions of the shock. Looking at total consumption (columns 1 and 2), we find consistently that the effect of the cash contraction was smaller in districts with a high penetration in electronic payments. Depending on the specification, districts with high penetration experienced a contraction in total consumption that is between 60% and 90% smaller than in low penetration areas.[116]

These results show that the cash contraction had a negative effect on individual consumption. However, the negative effects were somehow limited to the most acute period of the Demonetization. Furthermore, the cut was larger for unnecessary goods, like recreational expenses, and much more limited for food expense. Building on these patterns, we also show that the presence of a developed electronic payment infrastructure in a local market explains part of the variation in the response to the shock in the local market. This evidence suggests that – while electronic money was not sufficient to completely shield the economy from the contraction – its presence may have played a role in limiting the costs of the Demonetization.

Furthermore, this evidence is consistent with the interpretation of our specification as correctly capturing heterogeneity on the cash contraction. Consistent with this supply side interpretation, we find that our treatment predicts both lower consumption and higher use of electronic payments.

D.3 Placebo tests

In the body of the paper, we have also mentioned that the longer time series in the consumption data also allows us to run more detailed placebo tests on our treatment measure.

In general, before this test, one residual concern is that districts with high exposure to chest banks are regions that are particularly sensitive to business cycle fluctuations. The pre-trend analysis partially helps with this concern, but it cannot rule this out completely because it focuses on one specific point in time. Therefore, to bolster our identification further, we construct a large set of placebo tests, in which we repeat our main analysis centering it in periods in which there was no contraction in cash. In particular, to keep our approach general enough, we consider placebo shocks happening every month between February 2015 and February 2016. We then replicate our main specification, testing for the presence of a differential response across households in the wave of the placebo shock relative to the previous one.[117]

[115] The underlying assumption is that districts with a high number of ATMs per person will also be characterized by the highest concentration of debit cards and POS machines. We focus on ATM rather than directly on cards or POS, since we cannot directly measure the number of debit cards or POS machines at the district level, but only in aggregate.

[116] Table H.23 also examines the same effect across types of consumption. In particular, the access to electronic payments helped to reduce the impact of the shock in necessary consumption (columns 3 and 4),the impact in explaining the effect for unnecessary consumption was minimal (columns 5 and 6). This heterogeneity between types of consumption is consistent with both demand and supply mechanisms. On the one hand, consumers facing a scarce access to electronic payments may be more likely to allocate a larger share of their electronic money to necessary consumption. On the other hand, for necessary consumption – in particular food – consumers are more likely to face the option to trade with retailers that are larger in size (e.g. grocery chains) relative to unnecessary consumption (e.g. restaurants).

[117] In our main result, there is essentially no difference when we compare the effect on the previous wave - as in Figure H.18 - or the average of the previous three waves, like in Table H.20. Here we choose to compare to the previous wave because this

The results of this set of placebo tests are reported in Figure H.19. The general finding is that — in normal times — there is essentially no statistical difference in the change in total consumption between households in districts with different chest bank exposure. Together with the pre-trend analysis, this test excludes the concern that differential exposure to business cycles may explain our results. More broadly, this test provide new evidence on the validity of our empirical specification.

E State-dependence

In this Appendix, we go back to the state-dependence tests presented in Section 4.3 and discuss more in detail some of the potential identification shortcomings. In particular, discuss how the reflection problem Manski (1993) represents a general constraints to examining this type of problem, and also discuss how our approach helps overcoming this problem. We also present new empirical findings that help validate the quality of our setting. Lastly, we also discuss in detail the tests for state-dependence developed using the firm-level data.

E.1 Distance-to-the-hub

The idea of state-dependence in our model is that the response to the shock should not be uniform but it should crucially depend on the initial conditions: in particular, areas where the initial marginal benefit to join the platform is higher should see higher responses later. Because of this formulation, our starting point in studying state-dependence is to run the type of analysis suggested by the model: when this mechanism is important, we should find larger responses in those locations where the initial usage of the technology is more extensive. In fact, in our theoretical framework differences in initial conditions fully determine the marginal benefits of joining the platform. This relationship could be estimated using this equation:

$$X_{d,t} = \alpha_t + \alpha_d + \delta \left(I_d \times 1\{t \geq t_0\} \right) + \Gamma_t^{'} Y_d + \epsilon_{d,t} \tag{110}$$

where $X_{d,t}$ is a measure of the use of electronic payment in district d in month t, α_t and α_d are month and district fixed-effects, I_d measures initial adoption level in a district, and Y_d represents a vector of control at district-level. A model with state-dependence would predict that $\delta > 0$. Indeed, we find evidence that is consistent with this hypothesis (Appendix Table H.6).[118]

The key problem with this approach is that — as highlighted already in Manski (1993) — the estimate of the endogenous response due to network effect cannot be disentangled from the correlated and contextual effects. In other words, the estimated parameter $\hat{\delta}$ could capture both the effect of externalities as well as other contextual or correlated factors affecting both initial adoption and the post-Demonetization response.

To overcome these issues, we have introduced in the paper a test that examines whether the increase in adoption differs depending on the distance between a district and areas in which the usage of the electronic wallets was large in both absolute and relative terms prior to November (electronic payment hubs). In our specific setting, the idea is that retailers located closer to an electronic payment hub should be characterized by an higher marginal benefit to join the platform because of adoption externalities. For instance, consumers are more likely to travel across nearby locations, and therefore the vicinity to an electronic payment hub

allows us to go further back in time with the placebo.

[118] We specifically use two measures of pre-adoption in Appendix Table H.6: in odd columns, we use a dummy equal to one if the district has a positive amount transacted pre-Demonetization, and in even columns, we use a continuous variable equal to the log of the total amount transacted in the pre-period plus one.

should influence consumers' adoption of electronic payments which in turn should increase the incentive of local businesses to accept this form of payment. If this is correct, then we should find that the location should mediate the change in electronic payments' use around the Demonetization, since the benefit from externalities should be increasingly important as the scale of the technology expand.

Notice that the logic behind this test is similar to the typical argument used to motivate tests focused on the presence of "indirect network effects" (Rysman, 2019; Jullien et al., 2021): rather than exploiting variation in the size of the network to identify the endogenous response due to externalities, these approaches examine the relationship between two economic variables that should be related only under the assumption that network effects are sufficiently strong. As these authors argue, this alternative way to look at the problem helps overcoming the standard reflection problem and allows to think about identification in a more transparent way.

As described in the paper, we implement this approach by using a simple difference-in-difference model where we compare the usage of wallet technologies around the Demonetization period across districts that are differentially close to a digital wallet hub. We define hubs as those districts with high adoption in electronic payment pre-Demonetization in both absolute and relative term. We can also confirm that adoption levels in electronic payments around these areas tend to be higher also before the shock. With this model in mind, we estimate the following equation:

$$X_{d,t} = \alpha_{st} + \alpha_d + \delta\left(D_d \times 1\{t \geq t_0\}\right) + \Gamma_t^{'} Y_d + \epsilon_{d,t}$$

where $X_{d,t}$ is a measure of the use of electronic payment in district d in month t, α_{st} and α_d are month-by-state and district fixed-effects, D_d measures the minimal distance to one of the 5 electronic payments hubs, and Y_d represents a vector of control at district-level. Our prediction is that $\delta < 0$, which is that places further away should respond relatively less. As we described in Section 4.3, our results confirm our hypothesis, as we find that districts closer to an hub saw their adoption increase relatively more in the aftermath of the Demonetization.

The advantage of this approach is that it allows us to overcome the classical issues related to the reflection problem, since by construction the model does not rely on ex-ante differences in adoption. However, the interpretation of the test as evidence for state-dependence still requires a relative strong exclusion restriction. In particular, our approach would identify the role of externalities only if the distance from an electronic payment hub will affect adoption only because of its effect on adoption externalities. This concern could be interpreted within the traditional omitted variable bias framework. Assume we can decompose the error term $\epsilon_{d,t}$ as the sum of a purely idiosyncratic component $\xi_{d,t}$ and a vector of district-specific characteristics Z_d, such that each component of Z_t may affect electronic payment in a way that is captured by the vector Θ_t. In other words, $\epsilon_{d,t} = \Theta_t^{'} Z_d + \xi_{d,t}$, where $z_d^{(g)}$ and $\theta_d^{(g)}$ represents the component of vectors Z_d and Θ_t respectively. Within this framework, we can define $z_d^{(1)}$ to be the factor that captures the strength of adoption externalities in our model (i.e. the endogenous effect in Manski (1993)). This variable is unobservable to the econometrician, but — because of its relationship with distance — we can test for its importance in the data running the regression above. Given this framework, we can re-write the equation above as:

$$X_{d,t} = \alpha_{st} + \alpha_d + \delta\left(D_d \times 1\{t \geq t_0\}\right) + \Gamma_t^{'} Y_d + \Theta_t^{'} Z_d + \xi_{d,t}$$

In this framework, our identifying assumption is that there is no $z_d^{(g)}$ with $g \neq 1$ that: (a) is unobservable; (b) is correlated with distance D_d; and (c) has a significant effect on the use of electronic payment. In other words, the only $z_d^{(g)}$ that is allowed to be correlated with distance and also affects adoption is when $g = 1$ (i.e. distance only captures ex-ante differences in adoption externalities for electronic payments, which is the specific dimension we are trying to capture in the paper). An example of a $z_d^{(g)}$ that could pose a threat to our model is the presence of different propensity in adopting new technologies across areas that are closer to an electronic hub. This would be the case if areas closer to an electronic payment hub are systematically wealthier or are more familiar with better-tech products.

While this hypothesis is fundamentally untestable, we now provide a set of observations and tests that will help the reader to assess the plausibility of this assumption and provide some "boundaries" on the type of omitted factor that may pose a threat to our model. We discuss them here in order:

1. Controls: The first thing to point out is that the assumptions above needs to hold only conditional on the controls that are included in the analysis. As specified above, each specification always controls for a large array of district level characteristics that may affect the adoption of electronic payments. In particular, we control for variables that would capture the level of economic activity in the area, access to formal financial institutions, and distance to the state capital.[119] This last control is included because the distance to an electronic payment hub may systematically capture variation between more urban versus rural areas. Each control is included in the specification fully interacted with month fixed effect, essentially allowing each of these observable variables to flexibly affect the adoption patterns around the Demonetization. We also include state-by-month fixed-effects to make sure that the distance variable does not simply obtain variation from very heterogeneous part of the country, which may also be on a different trend in terms of adoption.

This first comment relates to the assumption (a) above (i.e. unobservability): we have included in the vector Y_d a large set of district characteristics that may be plausibly correlated with distance and also have an effect on electronic payment. Therefore, while we cannot fully control for all relevant variables, this first observation helps ruling out some obvious concerns. For instance, it rules out that differences in level of economic activity are what explains our results.

2. Pre-trends: Figure 7 in the paper documents absence of pre-trends in adoption: that is, the distance to an electronic payment hub only affects adoption after the Demonetization and not before. We find this result consistently across all the outcomes used in the analysis. Furthermore, the differences in the pre-period are not only statistically non-significant, but also small in size. Since the analysis also always include district fixed-effects, this lack of pre-trend implies that $z_d^{(g)}$needs to be a characteristic that did not affect the increase in adoption in the pre-period, despite being a significant force around adoption starting in November.

This property would be likely satisfied by our preferred interpretation: if distance captures variation in adoption externalities, we should expect this effect to become economically significant only after the Demonetization. This is because the strength of network effects depend on on the size of the network, and therefore we should expect to find significant effects only after a large shock (i.e. the Demonetization) that triggered a persistent structural break. However, a lot of other alternative interpretations are less likely to square with this result. In general, any factor $z_d^{(g)}$ that should also affect adoption growth before the Demonetization would not satisfy this assumption. For instance, if districts closer to an hub are characterized by more tech-savy citizens (and business owners), then these differences should be reflected in pre-shock period adoption in the form of a failure of parallel trends. In general, most alterantive explanations do not

[119]District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population, level of population and distance to state capital.

depend on the size of the network, and therefore should appear both before and after the shock if they are captured by the shock.

3. Observable Characteristics: Following the logic of the previous two tests, we next examine whether we find any systematic difference in ex-ante characteristics between districts that are closer to the payment hub. The idea is that a potential confounding factor $z_d^{(g)}$ with $g \neq 1$ has to be correlated with distance in order to be a concern for us. As a result, it is plausible to expect that this factor should in part reflect on some other characteristic of the district. To examine this issue, we test for differences in distance from the payment hubs affect ex-ante observable characteristics. The results are presented in Appendix Table H.7. Conditional on the distance to the state capital and state fixed-effects, we find that distance to an hub is not correlated with ex-ante differences in characterstics. In particular, districts are also on average similar across their banking characteristics — they have same level of deposits in quarter before the Demonetization, have similar access to ATM machies as well as banking and credit facilities. The districts are also balanced on population, and socio-economic characteristics (e.g. literacy rate). Importantly, districts farther from the payment hubs do not differ in their exposure to the shock. This is an important point as it implies that our estimate of δ is not mechanically picking up differential adoption responses in far districts to low exposure to the Demonetization. Lastly, the districts have similar ownership of alternate payment technologies including credit card, cellphone ownership and banked population. Altogether, this evidence suggests that any potential confounding factor $z_d^{(g)}$(on top of the potential conditions discussed above) should also not affect ex-ante characteristics of the district.

4. Placebo: To complement the previous analyses and further examine which type of $z_d^{(g)}$ may be a concern for us, we also propose a series of placebo tests by examining the adoption of other technologies around the same period. The theoretical foundation for this test is the following: most unobservable factors $z_d^{(g)}$ with $g \neq 1$ that will affect the adoption in electronic payments should also affect the adoption of other technologies that are not electronic payments, but share similar adoption hurdles. For instance, any $z_d^{(g)}$ that captures some propensity to adopt new technologies in the local market should affect electronic payments but also impact households or retailers decision in the use of other fintech products or other technologies. Notice that this would not be the case for $z_d^{(1)}$, since this factor captures adoption externalities relative to the electronic payment network, which are not always relevant for other technologies.

To implement this test, we run the same distance analysis using specification (16) and present the dynamic results graphically. The reason why we particularly want to stress the importance of the dynamic specification here is because this test is not only looking for changes in adoption around the Demonetization, but we want more broadly to examine whether distance plays any role in explaining adoption dynamics during the period considered. In other words, this test is interested in determining whether: (a) distance affects adoption dynamics in general (i.e. areas closer/further experience differential growth); (b) and particularly whether distance affect differential changes in adoption around the Demonetization.

To examine this issue, we consider adoption of three alternative technologies: banks accounts, mobile phones, and fintech loans. While none of these benchmark variables are perfect, our argument is that these technologies are likely to be affected by similar factors that are relevant for adoption as electronic payments. First, we measure the adoption of mobile phones and bank accounts by households. Information on mobile phones and bank accounts is constructed using the CMIE household survey that is already employed to conduct the consumption analysis in the paper (Appendix D). Second, we measure district demand for fintech loans (i.e. loan applications), from one of the largest fintech lenders in India. This technology is more directly comparable in terms of penetration to electronic payments: it is a two-sided platform and local market forces that affect adoption of electronic payments will arguably also affect the adoption of this technology.

We explore all three dimensions since we think there are benefits and weakness with all three measures. From a conceptual standpoint, it is not immediately clear which of the three variables is better to conduct this placebo analysis. On the one hand, the data on fintech loans is more likely to be affected by the same type of consumers' preferences and adoption frictions as our electronic wallet data, as they both are fintech products. On the other hand, however, mobile phone and bank accounts can also be interesting to study, in particular since their adoption may be directly affected by the Demonetization. Similarly, from a data standpoint, it is not clear which of these technologies provide the best benchmark, since each of these technologies provides a snapshot of a technology that is at a very different phase of its life-cycle. While the use of mobile phones and bank accounts was extremely widespread before the Demonetization, the penetration of fintech loan product was relatively low in the early part of our sample. Altogether, we think that examining all three is the best and more transparent approach.

Across all three measures, we consistently fail to identify any systematic relationship between the distance from an electronic hub and the adoption of any of these technologies. The results are provided in the three panels of Figure H.10. While results are sometimes a bit noisy on a month-by-month comparison, the overall pattern clearly excludes any consistent difference in trends either before or after the Demonetization. Furthermore, we also do not find any change in the adoption trends of these technologies around the Demonetization event. Therefore, the within-state distance to a hub is not only unrelated to an overall trend in each of the three technologies, but it also does not appear to be correlated to any significant change in trend around November 2016.

Altogether, this set of results help us to further characterize the type of $z_d^{(g)}$ that could be a concern for our specification. In particular, this analysis suggests that $z_d^{(g)}$needs to be a factor that — on top of satisfying all the other conditions expressed above — also does not affect adoption of other technologies similar to electronic payments during the same period. In particular, these factors do not explain neither the changes of these technologies before the Demonetization nor a change in adoption around it. As argued above, this set of tests help us rule out a large class of potential confounding factors that are likely to impact both the adoption of electronic payments as well as other comparable technologies.

5. Sensitivity: Leveraging on the same data collected in the previous step, we also provide a sensitivity test where we examine whether our main distance result (i.e. specification 16) changes when we control for a variable that is likely to be correlated with $z_d^{(g)}$. In the previous test (4), we have shown that distance does not affect changes in adoption of other fintech products around the Demonetization. As a last test, we use the data on fintech loans to create a proxy for fintech familiarity in a district and use this variable as a control in the analysis. Specifically, we use the amount of fintech loan demand in the district per capita. While this test is related to the one in (4), the logic behind the test is slightly different. Previously, we wanted to check whether the same patterns could be identified on other technologies, under the assumption that we should be able to replicate this effect if our results were driven by some omitted factor that is also relevant for other technologies and correlated with distance. In this test, we take a more agnostic approach and simply examine the sensitivity of our main results when we allow districts with different ex-ante levels of fintech familiarity to have differential effect on electronic payments. To the extent that there exists a $z_d^{(g)}$

that can be a confounding factor here assuming this is correlated with our proxy of fintech familiarity, we should see our main coefficient of interest δ to change.

The results are reported in Table H.10. In particular, in odd columns we include our baseline results and even columns we include our results that also add our fintech familiarity by month controls. Across all our main outcomes of interest, we consistently find no economically and statistically significant changes in our main coefficient of interest. Building on the previous discussion, we should have found δ to change by some margin if our results where driven by an omitted variable that is somehow correlated with differences in fintech familiarity.

6. Heterogeneity: As a last result supporting state-dependence, we examine how this mechanism should interact with our main analysis. In principle, the state-dependence mechanism should reinforce the effect of the cash contraction (i.e., the effect of the shock should be larger where the initial marginal benefit to join is sufficiently large ex-ante). Our evidence confirms this hypothesis: in Appendix Table H.8 we show that the impact of a cash contraction is statistically stronger for areas that are located closer to a payment hub. This evidence is consistent with the model — which would predict a larger responses from areas with higher ex-ante marginal benefit to join the platform — and therefore it supports our idea that state-dependence is an economically important mechanism in our setting.

Altogether, this set of analyses helps us assess the plausibility of our interpretation of the distance test as evidence consistent with state-dependence. Our key identification assumption is that (conditional on the various controls) distance from an electronic payment hub will affect adoption only because of its effect on adoption externalities. The presence of some omitted factor that drives adoption of electronic payment and also correlated with distance is the main threat to this hypothesis (i.e. $z_d^{(g)}$with $g \neq 1$). While our tests cannot rule out this concern completely, they help address some of the alternative interpretations. In particular, conditional on the controls in the analyses, these omitted factors would have to: (a) be correlated with distance to an electronic hub; (b) affect the rise in electronic payments after the Demonetization, but not before; (c) not affect the adoption of other technologies; (d) be uncorrelated with fintech familiarity.

E.2 Firm-Level Tests

On top of the tests based on the distance-to-the-hub, we also find support for state-dependence using firm-level analyses.

In the model, analogously to the district-level prediction, state-dependence implies a positive relationship between a firm's use of the technology and the overall use by other firms in the same area. Using firm-level data, we can directly test this prediction, following an approach that is consistent with the empirical literature on spillovers (e.g. Munshi 2004, Goolsbee and Klenow 2002). Furthermore, the use of firm-level data allows us to control directly for several dimensions of heterogeneity that may explain adoption decisions for reasons unrelated to externalities.

For each firm, we measure the total use of the technology by other companies located in the same geographical area and operate in the same industry. We choose this reference group because we believe that complementarities should be strongest among firms in the same area and industry. For instance, we expect to find the largest overlap in customers for companies within the same area and industry, as well as the largest spillovers in learning about the value of the technology. [120] In particular, we estimate:

$$x_{i,p,k,t} = \alpha_i + \alpha_{p,t} + \alpha_{k,t} + \rho x_{i,p,k,t-1} + \gamma X_{p,k,t-1} + \epsilon_{i,p,k,t}. \tag{111}$$

Here $x_{i,p,k,t}$ is a measure of technology choice by firm i in industry k and pincode p at time t (where t is a week in the period May 2016-June 2017).[121] For instance, this measure could be a dummy for whether the firm used the platform, or it could be the amount of activity of the firm on the platform.[122] The variable $X_{p,k,t-1}$ is a measure of adoption by other firms in the same pincode and the same industry during the previous week. To be consistent, we measure $X_{p,k,t-1}$ using the same variable we used as the outcome, summing that dimension across all firms in the same pincode and industry, and always excluding the firm itself.

Results reported in Table H.12 provide evidence consistent with state-dependence. Across several specifications, we find that a higher volume of electronic transactions by firms in the same reference group strongly predicts more transactions for the firm itself in the following week. For instance, in our baseline we have that a one-standard-deviation increase in transactions by firms in the reference group leads to a 40% increase in the amount of transactions for the firm, which corresponds to 18% of the standard-deviation of the outcome variable. The same results hold — with similar magnitude — when we look at the number of transactions or at whether the firm was active on the platform.

Overall, the main concern in this analysis is that past decisions by firms in the reference group may correlate with an individual firm's behavior because of unobservable heterogeneity across firms which are unrelated to the strength of complementarities — the reflection problem. To assuage this problem, we show that results still hold once we augment the baseline with firm fixed-effects (column 2), pincode-by-week fixed-effects (column 3), and industry-by-week fixed-effects (column 5) altogether. Relative to the baseline specification specification (column 1), the addition of these fixed-effects will allow us to keep constant in the model any characteristics of the area — even to the extent that these characteristics have a differential effect over time — and also adjust the estimates for changes in adoption rates in the same industry.[123]

We conclude by repeating the same analysis as before, but allowing for month-specific parameters for each of our outcomes (Figure H.14).[124] Across the three outcomes, there are two key findings. First, the positive effect documented before is always present in the data, both before and after the policy shock. This is reassuring, since the state-dependence induced by complementarities is not a function of the shock but a feature of technology choices in any scenario. Second, the effect of adoption in the reference group is much higher in the months of the Demonetization, relative to the preceding and succeeding months.

[120] Our results also hold when using alternative definitions of the reference group. For instance, in Table H.13 in the Appendix we define the relevant market as any firm in the same location (pincode), irrespective of the industry.

[121] We use pincode to identify firms' locations because we want to use the narrowest definition of location that is available in the data. Our main results also hold using districts (Table H.14 in Appendix).

[122] We classify firms into 14 broad industries: Food and Groceries (14%), Clothing (10%), Cosmetics (2%), Appliances (8%), Restaurants (12%), Recreation (2%), Bills and Rent (1%), Transportation (13%), Communication (12%), Education (3%), Health (7%), Services (4%), Jewellery (1%) and Others (11%).

[123] The inclusion of individual fixed-effects in a dynamic model may bias the main parameters in the model, as first discussed in Nickell (1981). However, there are two important things to highlight about our application. First, the presence of fixed-effect is not necessary to obtain the desired result, since we still find the same effect without any fixed-effects (column 1). Second, the Nickell bias is a feature of models characterized by short panels, as the bias converges to zero as $T-1$ increases, where T is the time-dimension in the panel. In our case, T is relatively large - data is at weekly level and the time span is almost a year - and therefore the bias will be small in magnitude. In particular, since our main prediction is on the direction of the relationship rather than on the exact magnitude, this issue will not affect the conclusion of this study.

[124] In fact, the model suggests that the effect estimated should actually be different across time. In particular, using the simulated data from the model, we can show that the importance of the adoption by other firms is particularly large in the shock period.

In general, this firm-level tests confirm the takeaways from the analyses using the distance-to-the-hub: a local market initial conditions in terms of technology adoption matters for the propagation of the technology. In general, firms or districts that were more likely to face high marginal benefits to use the technology experienced a larger use of the technology ex-post. While both empirical models have some limitations, their combined evidence provides a relatively strong test for the state-dependence prediction described in the model.

F Learning vs. Network Effects

As we discussed in the paper, the presence of externalities in adoption in our framework can arise for a variety of reasons. While quantifying the relative importance of these different channels is outside the scope of the paper, this Appendix Section aims to present a variety of tests that can help the reader understanding the relative importance of the different channels. This section provides a more extensive discussion of what it is already presented in Section 4.5 in the body of the paper.

There are two main channels that may generate externality in adoption between retailers in our context. First, complementarity in adoption between retailers could be generated by network effects arising from the two-sided nature of the payment technology. For instance, in our context, the adoption by a retailer increases the value of adopting the same technology by other retailers because it makes the technology more valuable for consumers. This feedback-loop from the two sides of the market generates a positive externality in adoption. (Appendix B.5 provides a two-sided, micro-founded model consistent with this mechanism, and shows that it is isomorphic to our baseline model.) Second, externalities in adoption may also be induced by retailers learning about the technology, in a context where the relative benefits of technologies is uncertain: as more individuals use the technology in a local market, information about the existence and benefits of technology will be more widely available to retailers (either through direct observation or communication) which in turn should increase their likelihood to adopt.[125] While the two mechanisms could generate similar observational effects, they may have different policy implications.

To be clear, a third mechanism that may generate externality in adoption is learning between consumers. While we discuss this channel later, we also want to point out now that this alternative mechanism is different from the other two because its ability to explain the data is predicated on the assumption that network effects between consumers and retailers is an economically important mechanism. In fact, in order to affect retailers, learning between consumers requires the presence of a feedback-loop between the two sides of the market, as implied by the traditional network effect channel discussed earlier.

Separating learning from network externalities is a notoriously challenging task. However, we present here four separate tests that highlight the importance of network externalities in explaining our findings.

1. Use by pre-adopters. To start, we look at this issue by focusing on firms that were already using electronic payments before November 2016 ("pre-adopters") and had little to learn about the benefits of technology arising from the shock.[126] This idea follows the conceptual framework in Fafchamps et al. (2021): if the presence of complementarity in adoption between different retailers come from the presence of learning

[125] Learning may be important because there is some intrinsic level of ex-ante uncertainty about the nature of the technology. In this context, having more users in the platform may provide a signal about the quality of the technology, therefore resolving some of the initial uncertainty and increasing the value of adoption.

[126] We define as a pre-adopter those merchants that have conducted at least Rs. 50 of transactions by October 2016.

between each other, then their ability to influence each other should be null when a retailer has already adopted the technology.

To be clear, as in Fafchamps et al. (2021), this implication follows from a very specific assumption about the nature of learning. In particular, we assume that having adopted the technology allows the user to learn most of what is important to determine future adoption. Given the relatively simplicity of the technology, we think that this assumption fits well our setting. However, it is also important to highlight this working assumption, and clarify how this may not be a good approximation for all problems.

An implication of this argument is that we should not see any persistent increase in the use of electronic payment for pre-adopters around the Demonetization if the only source of externality in adoption is learning: as cash comes back into the system, their use of electronic payment should go back to the pre-shock level. This is obviously not true if traditional network externalities are somehow important: in this case, the cash crunch should have persistent effect, because of the feedback-loop between consumers and firms.

The data confirms this second hypothesis: pre-adopters saw a persistent increase in the use of electronic payments in the period considered. In aggregate, the firms that were pre-adopters (i.e. users in October 2016) experienced a substantial increase of about 100% in number of transactions between October 2016 to May 2017. Using our analysis exploiting variation across districts, we also find a large persistent effect of the shock in this sub-sample of firms: in Table H.15, we now conduct our main analysis using firm-level activity for this set of pre-adopters.[127] On top of finding that these firms also increased the use of electronic payment on average after the Demonetization (column 1), we also find that the effect is still large and significant in both the short- and long-run (column 2). In fact, in this specification, we estimate a separate effect for short-run (i.e. November 2016 to January 2017) and long-run (i.e. February 2017 onwards), and find that the long-run effect is still very large and statistically similar to the short-run effect. This empirical observation would be inconsistent with the idea that the persistence is mostly driven by a learning process.

2. Use by early-adopters. As a second test, we follow the same logic as the previous one but now look at a different subset of users in our data: those that adopted electronic payments in the short run during the Demonetization period. For this group we study the growth in their use of electronic payment during Spring 2017. Our argument is that the degree of connection between the regional exposure to the Demonetization and the long-term transaction growth for these early adopters depends on the mechanism that generates externalities.

Consider the case when the externality in adoption between retailers is completely determined by retailers learning about the technology, either by observing adoption decision of other retailers or by learning through interactions. In this case, we should expect no relationship between the growth in transactions during Spring 2017 and the exposure to the shock in November 2016 (or potentially a negative relationship if we think that the adoption process is characterized by mean reversion).

This result follows from two observations. First, in Spring 2017 cash crunch has already dissipated and therefore the November shock should only affect the use of electronic payment indirectly, through its impact on the aggregate use of mobile wallets in a local market. Second, the businesses we are considering have already adopted by January 2017, and therefore they have already learnt about the technology before Spring 2017. This implies that other contextual factors —for example, the adoption decision of the other firms — should not be relevant anymore, if externalities only operate through learning.

[127] To be clear, this analysis is conducted using the firm-level data on the sample of pre-adopters rather than the aggregated data at district level. Therefore, while the specification employed and the variable construction are identical to our main district-level analyses, we can now also include firm fixed-effects.

Instead, if network effects are a primary determinant of externalities, we should expect a positive relationship between the size of the shock and the growth in number of transactions in Spring 2017. The intuition behind this result is straightforward: when network effects are important, the size of the network should positively affect the use of electronic payments in the long run for firms that have already adopted. The testable prediction is that if the network effects are the key driver of our results, then the regional exposure to the Demonetization should affect the use of electronic payment in Spring 2017 also for firms that have already adopted the technology. Also in this case, the logic behind our test would be consistent with the conceptual framework presented in Fafchamps et al. (2021).

We test this idea by focusing on firms that have adopted in the Demonetization period (i.e. November 2016 to January 2017) and regress the firm-level growth rate in the number of transactions on mobile wallet for these firms during Spring 2017 (i.e. the growth rate between March and June 2017) on our proxy of the cash shock exposure.[128] The results are reported in the two panels of Figure H.13. In general, we find a statistically positive relationship between the long-run growth rate and the shock (panel a), which also survives when we control for the number of transactions conducted during the Demonetization period (panel b). This evidence appears consistent with a model where network effects play an important role in determining externalities in adoption.

These two sets of tests on the long-run growth for early- and pre-adopters can be easily rationalized in a context where traditional network effects are relevant, while canonical models of learning alone would fall short in explaining these patterns. However, this does not imply that learning from retailers is not a relevant aspect, but rather that this force cannot entirely explain our results alone.

3. Heterogeneity by (proxies for) social learning. As a third test, we examine whether the response to the shock is different depending on how easy is for local agents to collect information. The discussion has so far focused specifically on learning by retailers, i.e. retailers learn about the technology from other retailers or consumers. As discussed earlier, it is particularly important to examine this mechanism because it can generate externalities in adoption also without the presence of any feedback-loop between the two sides of the market. However, learning could also happen among consumers. This alternative mechanism is intrinsically nested within the traditional network effects since learning on the consumer side will only be relevant for retailer-adoption when network effects across the two markets are relevant. As a result, in our case the two mechanisms are hard to separate, both empirically and conceptually.

To examine learning more broadly, we test whether there is a stronger response to the shock in areas where learning is easier. We consider two proxies for consumer learning in a region. First, we use the degree of language concentration.[129] Second, we also examine the extent to which the population of a district is connected to other people from the same district on Facebook (Bailey et al., 2018).[130]

In general, if learning is a first-order mechanism, we should expect to find a stronger increase in adoption

[128] Specifically, we estimate equation $\Delta Y_{id} = \beta$ Exposure$_d + \Gamma X_d + \epsilon_d$, where ΔY_{id} is the growth rate in $Y_{id,t}$ between $t =$ March 2017 and $t =$ June 2017. $Y_{id,t}$ are the number of transactions in month t for firm i, located in district d. The sample considered is all firms that have had a positive amount of transactions for the first time between November 2016 to January 2017, and are still receiving payments as of March 2017.

[129] We define language concentration in a district as: Language Concentration$_d = 1 - \sum_l s_{dl}^2$ where s_{dl} is the share of district d population speaking language l. We obtain the information on language distribution among population using Census of India, 2011. We also standardize the measure to have mean 0 and standard deviation of 1.

[130] We use Facebook's Social Connectedness Index (SCI) to measure the degree of digital connectedness. SCI measures the number of Facebook friend connections between districts i and j divided by the product of numbers of Facebook users in the two districts. We also standardize the measure to have mean 0 and standard deviation of 1. Given that SCI uses GADM's 2014 border for Indian districts, we aggregate the intra-district SCI value to 2001 Census border assuming that number of users is proportional to the total population at each district for the SCI denominator (i.e. population weights).

in districts where learning is easier, like districts with more homogeneous languages or where individuals are more connected with each other through social networks. To start, we examine this question using our district-level data. We test whether the new adopters is influenced by our two measures of the ease of learning (Appendix Table H.16).[131] Both measures reject the hypothesis that places where learning is easier experienced a large increase in adoption. As an aside, we also replicate the same effect looking at the sample of pre-adopters discussed earlier. Since these are all retailers that have adopted, we now focus on an intensive margin of the use of electronic payment (i.e. amount transacted). As we show in columns 3-6 in the Table H.15, we do not find any evidence that areas where social learning is easier respond differently. These effects appear at odds with what one may expect if learning is the driving force behind our results.

4. Results from survey. As a fourth (and final) test, we also ran a survey of small firms and consumers in India that adopted electronic payment during the Demonetization. The objective of the survey is to try to elicit information on the factors that the respondents consider most important in deciding whether to adopt electronic payments. We ran the survey using MTurk, a platform that is frequently used to run experiment and surveys online and that also allows to run studies targeting adult individuals living in India. The survey is presented to respondents as a general study on the use of electronic payment during the Demonetization. Individuals are asked questions that allow us to determine whether the individuals have adopted any form of electronic payment in the post-Demonetization period, and whether they identify themselves as retailers or consumers.

The key question asks them about the reasons that pushed them to adopt electronic payment during the Demonetization. Since we initially ask them whether they are small business owner, we tailor the language of specific question to either "consumers" or "retailers." The survey provides them with three pre-set options to pick (presented in random order to the respondents) plus they have the opportunity to add another open response. The first option aims to measure the direct impact of cash crunch in affecting their decision to use electronic payments ("New Cash was hard to find, and therefore I had to find other ways to pay for things"). We have this option in the survey because it helps us generate a benchmark for the importance of the other mechanism. A second option proxies the traditional network externalities that would be generated in a two-sided market. The exact wording of the question is slightly different for consumers and business owners, since in each case we motivate the adoption of electronic payment with an increase in the use of electronic payment on the other side of the market. For instance, for consumers, we motivate their adoption of electronic payment as the result of a change in the use of electronic payment by the shops they commonly use ("Because the shops where I buy things started accepting non-cash payments, it was better for me to use this option"). For business owners, we instead frame it about their customers ("Since my customers had started using non-cash payments, it was better for me to offer this option."). A third option instead tries to proxy for learning, essentially saying that the main reason for adoption was that they have learned about the technology from family and friends. We allow respondents to pick multiple options: for instance, people can select the cash crunch and also the learning response.

Our initial sample is made up 664 responses.[132] However, we immediately drop responses that are not

[131] Since the measures of language concentration or SCI are unavailable for some districts that were part of our main sample used in Table 2, in Column 1 of Appendix Table H.16 we re-conduct our main analysis on the sub-sample and show that our result also replicates well on this sample.

[132] Participation was clearly voluntary and compensated with $1 for a study taking less than 5 minutes. Because of concerns regarding data quality, we have included several filters in the initial Mturk setting. For instance, we clearly exclude repeated participants and only allow participants with high HIT approval ($> 90\%$) and number of HIT approved between 50 and 10,000. The survey was ran between December 2021 and March 2022.

complete (i.e. the person did not finish filling the poll) or that failed our attention checks, dropping the same to about 544 respondents.[133] We also keep only those that (1) used electronic payment over the past 5 years; (2) adopted some form of electronic payments during the Demonetization. After this second filter, 432 unique responses remain. While this sample is not big in absolute terms, we think it can provide useful information for our answer.[134]

We present the result of the survey in Table H.17. We find that the "network effect" option — which highlights the adoption by their shops or customer — is the most selected, as 75% of respondents chose it. This is high in absolute terms, but it is also high relative to the number of individuals that instead identified in the direct effect of cash the main reason for adopting electronic payments (56%). The role of learning appears relevant, but less important than the other two options: only 44% claims that the adoption of electronic payment was affected by having learned about the technology from friends and family. Almost no one has selected the other option.[135]

Altogether, this evidence must be considered with the usual caveats regarding the use of surveys. However, the picture that comes out of this exercise is surprisingly consistent with our general narrative. Both learning and network effects appear to be relevant to understand the adoption of electronic payment during the Demonetization. Within this context, however, network effects appear to be a more significant force, as individuals point out on the increase in the use of electronic payment on the other side of the market as an important factor in affecting their decision.

Before concluding, it is also important to highlight two weakness of this survey. First, as we pointed out before, the survey was run about five years after the Demonetization. Second, since the survey was ran using Mturk, the sample is likely to be biased towards a subset of the Indian population that is more likely to use internet services. Indeed, Table H.18 provides evidence consistent with this concern: our survey coverage is relatively more representative of younger Indians living in the Southern part of the country. About 96% of our respondents are 50 or below, compared to only 74% in the overall population.[136] Furthermore, about two-third of our survey respondents are from one of the Southern states, compared to about one-fourth of the total population. We believe that these differences capture the higher propensity in using internet services - such as Mturk - among younger adults and individuals living around the Bangalore area, which is the main technology hub in India.

To partially address these concerns, we present evidence consistent with our finding from the Demonetization survey ran by Financial Inclusion Insights India.[137] While we believe that our survey better targets the specific question we are after, this alternative work helps us address the two weaknesses discussed above. First, the survey was run right around the Demonetization. Second, according to the documentation, the survey covers a sample that should be representative of the adult Indian population.

[133] To screen for the presence of bots or potential individuals that do not pay any attention, we do two things. First, we have a set of two images (a picture of a dog and a picture of a banana) and we ask the person to pick what the figure shows. Conditional on paying any attention and not being a bot, answering this question is trivial, and indeed very few people make a mistake. Second, we ask them two questions about the Demonetization: who announced the policy, and what the respondents' experience in the aftermath of the Demonetization was like. The objective here is to also eliminate people that do not pay any attention (e.g. respond with random text) or have no knowledge of English.

[134] We also collect demographic information to make sure our sample had a reasonable coverage of the Indian population.

[135] We also repeat the same analyses by including more restrictive filter for excluding respondents. We also find similar findings when looking at only those that have adopted mobile wallets specifically. In general, the relative importance of the options does not change.

[136] Since the survey has been conducted only on individuals 18 or above (because of IRB concerns), the distribution is calculated on the same sub-group. Furthermore, to compare with the 2011 Census of India we had to do a few adjustment, since the data reporting brackets for age were slightly different. For instance, our youngest age group was 18 to 30, while with Census data we were only able to construct a 18 to 29 years old group.

[137] http://fii-website.staging.interactive.columnfivemedia.com/uploads/file/Effects%20of%20Demonetization%20on%20Financial%20Inclusion%20in%20India.pdf

Within this survey, we are particularly interested in the question reported in the Annex I.D.2 (page 33 of the report). This question focused on merchants that still did not accept cashless payments after the Demonetization and asked them for the motivation of this choice. A few non-mutually exclusive options are provided. Two of the options provided clearly point to a learning mechanism as the reason for not adopting: these options are (1) "Don't fully understand and/or unaware of this method"; (2) "Method is too difficult to use (self or customer)". Furthermore, another option points instead to network effects as the reason for not adopting. In particular, this option suggests that the lack of adoption is explained by the lack of demand from the other side of the market ("Customers don't demand this method"). Consistent with our findings, the network effects explanation appears relatively more important that the learning one. Each of the two learning explanations is selected by only about 60% of the respondents, while the lack of demand from the other side of the market is selected by 82% of the respondents. While this discussion does not address all the concerns with the survey, it does provides reassuring evidence that the importance of the network effects mechanism is likely not explained by the timing of the survey or the lack of representativeness of our sample.

To conclude, while this set of analyses cannot fundamentally decompose the relative importance of learning versus network effects in generating externalities in adoption, this analysis highlights that some degree of network externalities is probably necessary to rationalize all our findings. In other words, while learning between retailers could still be important in our setting, our evidence also suggests that the importance of this channel is likely less significant than the importance of network effects arising from the two-side market of this technology.[138]

G Estimation

G.1 Estimation method

Let Y and Z denote the dependent and independent variables in the system of equations (17); we first construct the OLS estimate of the data moments, $\hat{\Xi} = (Z'Z)^{-1}Z'Y$. We then estimate the variance-covariance matrix of $\hat{\Xi}$ using the bootstrap. Specifically, we let:

$$var\left(\hat{\Xi}\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\Xi}_b - \hat{\Xi}\right)'\left(\hat{\Xi}_b - \hat{\Xi}\right),$$

where $\hat{\Xi}_b$ is the estimate obtained in replication b of the bootstrap. We use $B = 100$ and sample with replacement district by district.

The point estimate for the $N_p \times 1$ vector of parameters Θ is obtained by solving:

$$\hat{\Theta} = \arg\min\left(\hat{\Xi} - \frac{1}{N_{sim}}\sum_{s=1}^{N_{sim}}\Xi_{sim}\left(\Theta;\gamma_s\right)\right)' W\left(\hat{\Xi} - \frac{1}{N_{sim}}\sum_{s=1}^{N_{sim}}\Xi_{sim}\left(\Theta;\gamma_s\right)\right).$$

In this objective, N_{sim} is the number of simulations, and $\Xi_{sim}(\Theta;\gamma_s)$ is the same vector of moments as above, estimated using data produced by simulation s. We use $N_{sim} = 20$ simulations, in keeping with the

[138] Another result that appears inconsistent with a traditional model of learning is the presence of reversal. In common feature of learning in canonical models is that learning cannot be undone (at least within a few months). One implication of this feature is that there should be more limited reversal after the shock if learning were the key source of complementarities. Indeed, the data seem at odds with this scenario, since slightly more than a quarter of our district-month pairs experienced some negative growth after the Demonetization in our data.

recommendations of Michaelides and Ng (2000). Each simulation has the same size as the panel data; data is sampled monthly from model simulations. We simulate data with a burn-in period of 10 years for each district. Additionally, γ_s is a vector of random disturbances for simulation s, which we keep constant across values of Θ for which the objective is evaluated. We use Matlab's patternsearch routine to minimize the objective, with 20 randomly drawn starting points for Θ.

Following the litterature (Pakes and Pollard, 1989; Rust, 1994; Hennessy and Whited, 2005, 2007; Taylor, 2010), we use the optimal weighting matrix:

$$W = \frac{1}{N_m} var\left(\hat{\Xi}\right)^{-1}.$$

The variance-covariance matrix for $\hat{\Theta}$, the vector of estimated parameters, is obtained as:

$$\Omega = \left(1 + \frac{1}{N_{sim}}\right)\left\{\left(\frac{\partial G}{\partial \Theta}(\hat{\Theta})\right)' W \left(\frac{\partial G}{\partial \Theta}(\hat{\Theta})\right)\right\}^{-1},$$

with:

$$G(\Theta) \equiv \hat{\Xi} - \frac{1}{N_{sim}} \sum_{s=1}^{N_{sim}} \Xi_{sim}\left(\Theta; \gamma_s\right).$$

We approximate the Jacobian of $G(.)$ using numerical differentiation. We also report the following test statistic for over-identifying restrictions:

$$J = \frac{N_{sim}}{1 + N_{sim}} G(\hat{\Theta})' \left(var\left(\hat{\Xi}\right)^{-1}\right) G(\hat{\Theta}),$$

which is distributed as a χ-squared with $N_m - N_p$ degrees of freedom under the null that the over-identifying restrictions hold. Additionally, we use 2000 simulations of the panel, with parameters set to $\hat{\Theta}$, to construct the standard errors and p-values reported in table 5.

In the data, we also re-normalize the Census retail counts so that at least $n \geq 0$ districts reach full adoption. Specifically, for all districts d, we define $X_{d,t} = \min(N_{d,t}/\bar{N}_d^{(n)}, 1)$, where $N_{d,t}$ is the number of adopters per district, and $\bar{N}_d^{(n)} = \frac{N_{d_n,t_0}}{N_{d_n}} N_d$, N_d is the Census count of retailers in district d in 2014, and d_n is a reference district. The reference district is defined as the district with the nth highest un-normalized maximum adoption rate, i.e. the nth highest value of $\max_t \frac{N_{d,t}}{N_d}$. We do this because it is unclear whether the Census counts properly measure the pool of potential adopters. We experimented with values ranging from $n = 0$ (no normalization) to $n = 10$ (the 10 highest-adoption districts reach full adoption). In all cases, we can reject the null of no complementarities, and estimates of the contribution of complementarities to the long-run change in adoption are largely unchanged, ranging from 40% to 65%. We use $n = 5$ in the estimation that follows.

G.2 Intuition for identification

Our main parameter of interest is the strength of complementarities, C. Consistent with our earlier discussion of the model, this parameter is primarily identified by the difference between the short and medium-run response of adoption to the shock, $\hat{\gamma}$.[139] Without adoption complementarities ($C = 0$), the short-run

[139] Here, we define "medium-run" as three months after the shock; by then, in the data, cash circulation had returned to pre-shock levels, and, in the model, the aggregate shock is more than 90% dissipated.

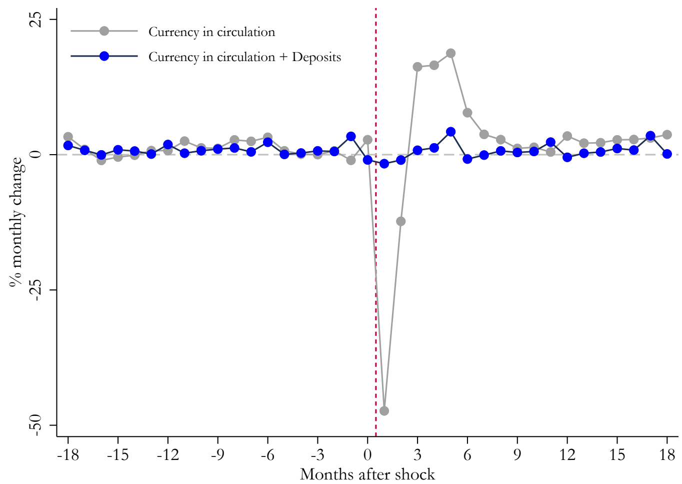
adoption wave triggered by the shock has no bearing on the adoption decision of firms further down the road. As a result, once the shock is dissipated, there should be no further adoption by new firms, consistent with Predictions 1b and 2b from Section 3. The model would then predict that $\hat{\gamma} = 0$. By contrast, when adoption complementarities are present ($C > 0$), the short-run adoption wave raises the value of future adoption for other firms, and so new firms continue adopting even once the shock is dissipated, leading to positive values of $\hat{\gamma}$. Additionally, as discussed earlier in the paper, the dependence of the response to the shock on initial conditions $(\hat{\delta}, \hat{\zeta})$ also helps pin down the strength of adoption complementarities.

The rate at which firms reset their technology choice, k, is identified using estimates of the between-district variance of the change in adoption, $\hat{\eta}$ and $\hat{\kappa}$. The medium-run variance, $\hat{\kappa}$, is particularly informative about k. As highlighted in our earlier discussion, if firms reset their technology quickly relative to the persistence of the shock (i.e. k is sufficiently high relative to θ), then all districts will rapidly converge to full adoption, thus leading to lower cross-sectional variance in adoption rates in the medium-run.

Finally, the size of the shock, S, is primarily identified by the short-run adoption caused by the shock, which is $\hat{\beta}$. Absent an aggregate shock, $\hat{\beta}$ is not statistically different from 0, and the magnitude of the coefficient increases with the size of the shock, independent of the existence of complementarities. The standard deviation of idiosyncratic innovations to districts, σ, is identified using the variance of residuals from the first equation in (17). The residual variation in adoption, after controlling for initial conditions, should be driven by district-level shocks. The rate of profits associated with the electronic payments technology when there is no adoption, M^e, is identified using the variance of within-district adoption rates. Even when there are no complementarities, a lower level of M^e is associated with shorter adoption spells, and therefore lower overall volatility of adoption rates.

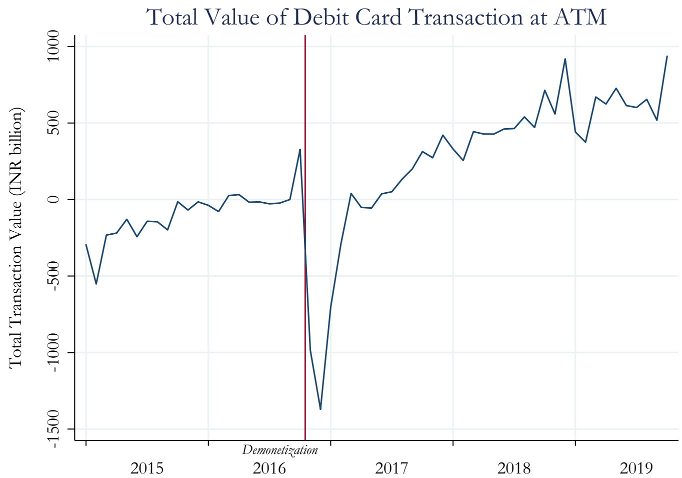
H Appendix figures and tables

Figure H.1: Nominal value of currency in circulation



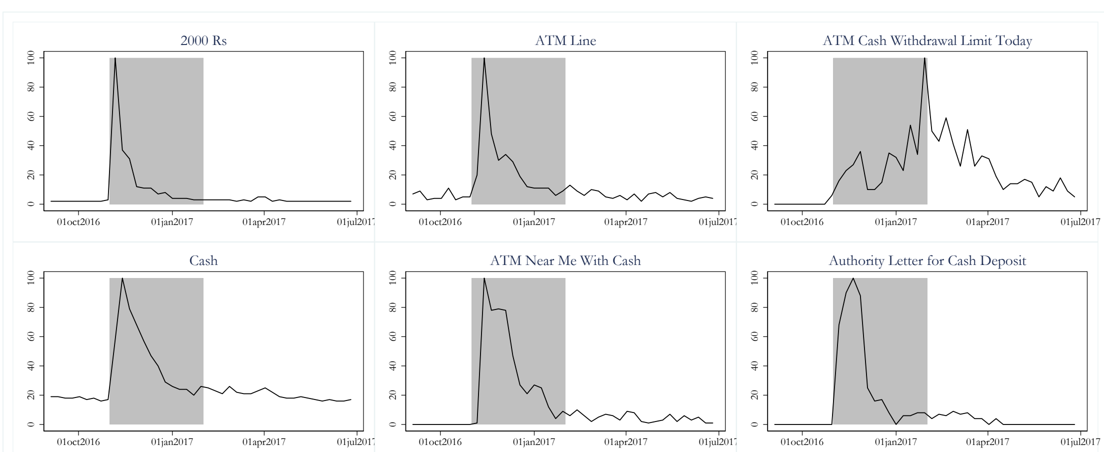
Notes: The figure shows the monthly change in the nominal value of currency in circulation (in grey) and the monthly change in the nominal value of the M1 money supply, the sum of currency plus bank deposits (in blue). Month 0 is the month of October 2016; the figures are end-of-month estimates. Source: Reserve Bank of India.

Figure H.2: Total value of ATM withdrawals in India (2015-2020)



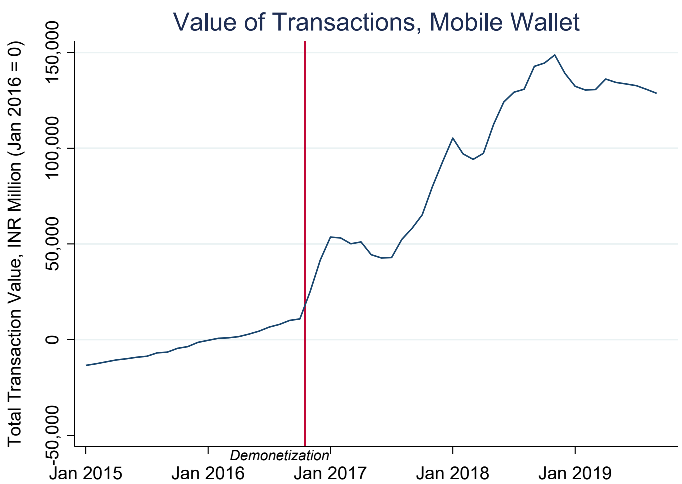
Notes: The figure reports the value of ATM transactions in India between 2015 and 2020. Specifically, this variable captures the amount of cash that is physically withdrawn from an ATM using a debit card during this period. We normalize at zero the level at October 2016. The vertical line is between October 2016 and November 2016. Source: Reserve Bank of India.

Figure H.3: Evidence from Google Search Trends



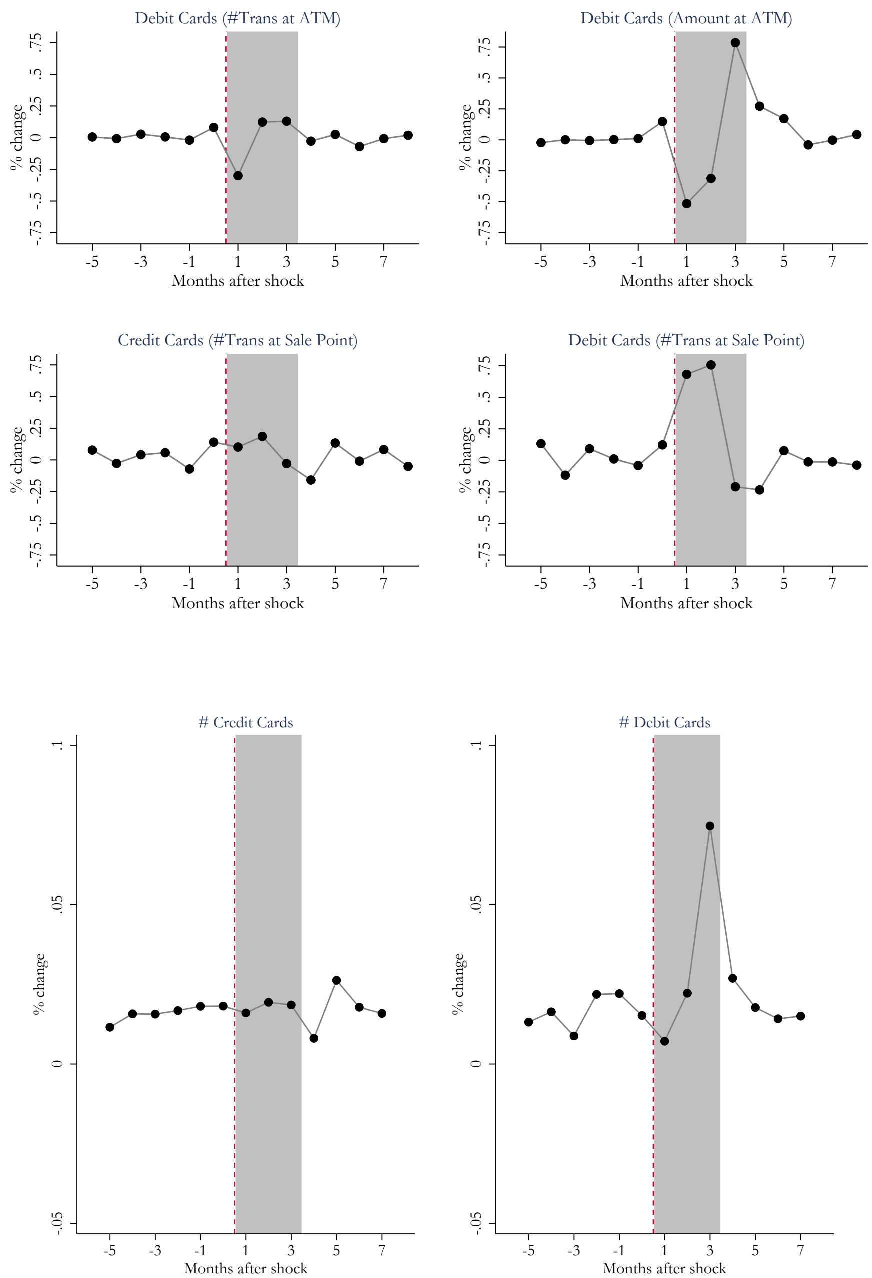
Notes: The figure reports the daily plot between September 2016 and July 2017 of Google searches for several key words that could be representative of public actions and information associated with the demonetization shocks. Data is obtained through Google Trends, and the index is normalized by Google to be 0 to 100, with a value of 100 assigned to the day with the maximum number of searches made for that topic. Source: Google Search Index.

Figure H.4: Total value of mobile wallet transactions in India (2015-2020)



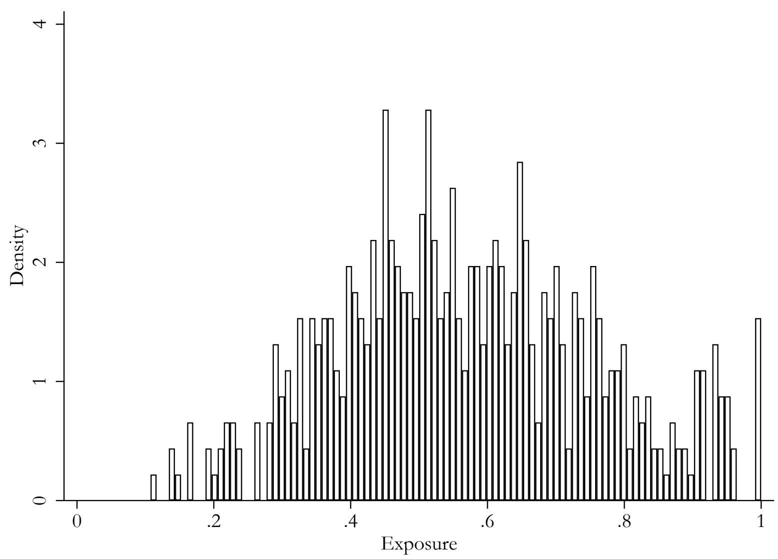
Notes: The figure reports the value of mobile wallet transactions in India between 2015 and 2020. We normalize at zero the level at January 2016, and smooth the series with a three month moving average. The vertical line is between October 2016 and November 2016. Source: Reserve Bank of India.

Figure H.5: Growth in Transactions for Traditional Electronic Payment Systems



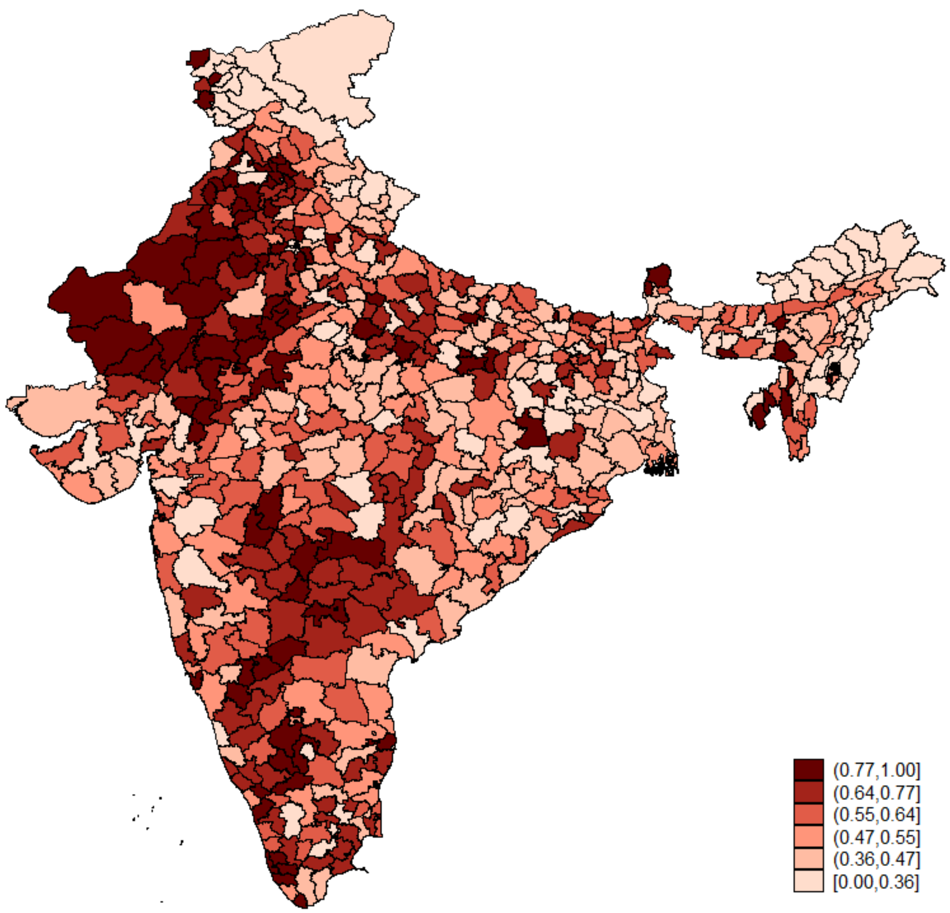
Notes: Monthly growth rates in transactions using credit and debit cards around Demonetization. The top four panels reports measures of use at the intensive margin, and the bottom two panels reports measures of adoption. All the data are monthly and aggregated at the national level. Months are on the horizontal axis, with October 2016 as month zero. Source: Reserve Bank of India.

Figure H.6: Distribution of Exposure$_d$ across districts



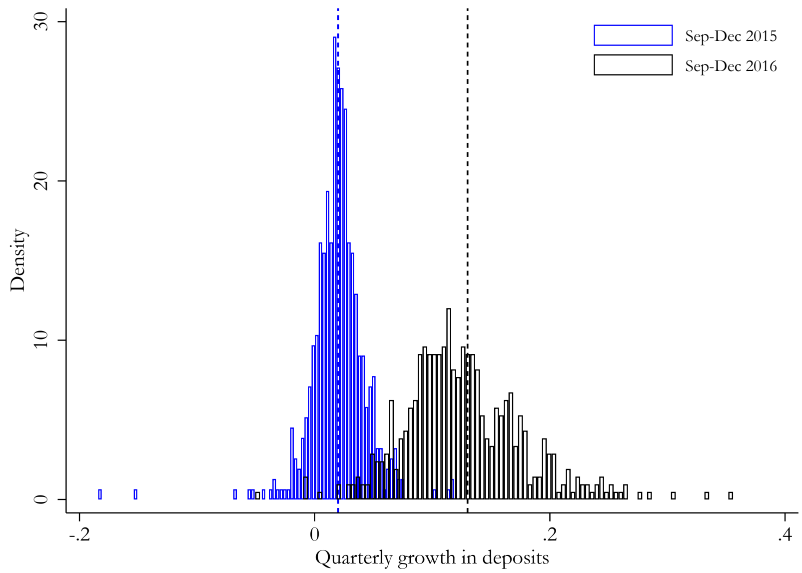
Notes: The figure shows the distribution of Exposure$_d$ (as described in Section 4) across Indian districts. Source: Reserve Bank of India.

Figure H.7: Map of the Distribution of Exposure$_d$



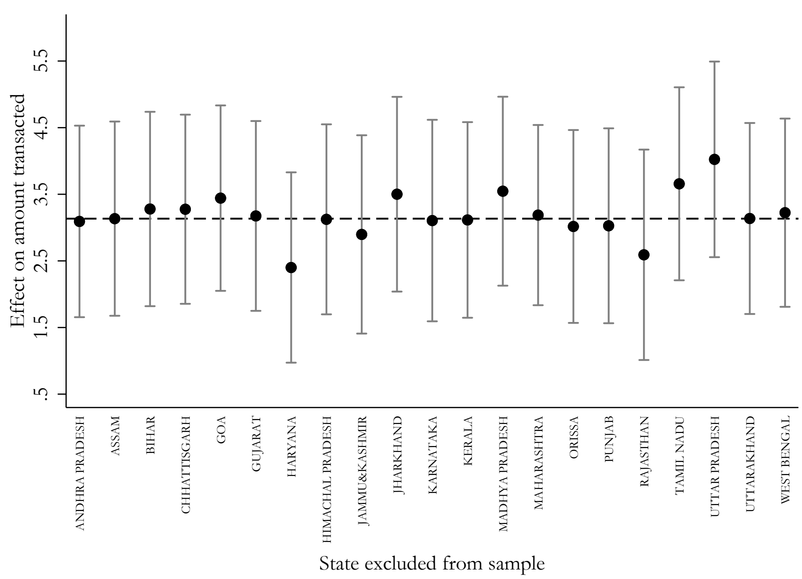
Notes: The figure maps the distribution of Exposure$_d$ (as described in Section 4) across Indian districts. Source: Reserve Bank of India

Figure H.8: Distribution of growth in deposits across districts



Notes: Distribution across deposits of the growth in total banking sector deposits from October to December during the year 2015 (blue) and 2016 (black). The vertical dashed lines represents the corresponding mean deposit growth for these years. Source: Reserve Bank of India.

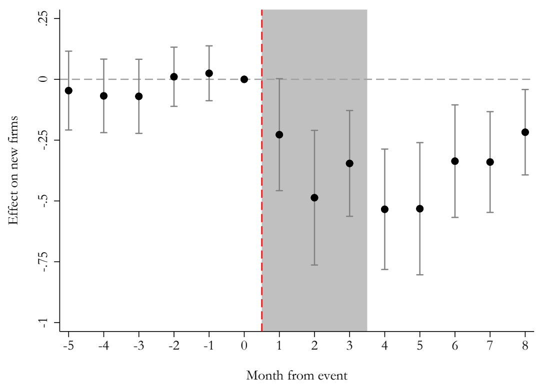
Figure H.9: Robustness: one-state out



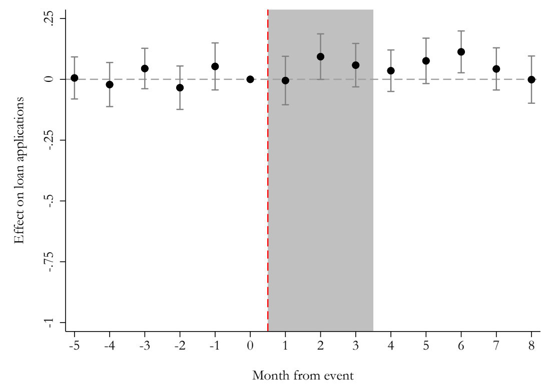
Notes: This figure reports a robustness in which we exclude from the main analysis one state at a time and we recalculate the main coefficient of interest. In particular, we consider the specification in which we look at amount of transactions as an outcome and we consider the coefficient on post multiplied to the chest exposure measure. Each bar reports the main coefficient for the specification excluding the state in the x-axis and the 95% confidence interval. The horizontal dashed line is the main coefficient from the main table of the paper, added for reference.

Figure H.10: District adoption dynamics across several technologies based on distance to electronic hub

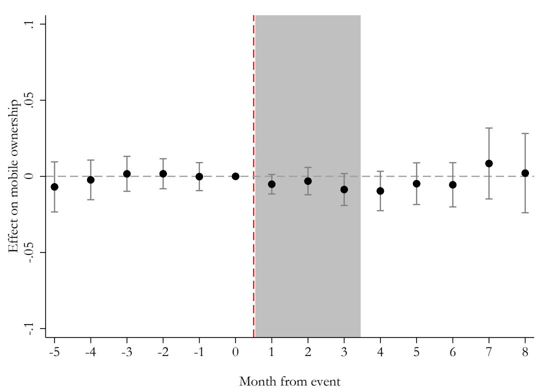
(a) Digital Wallet



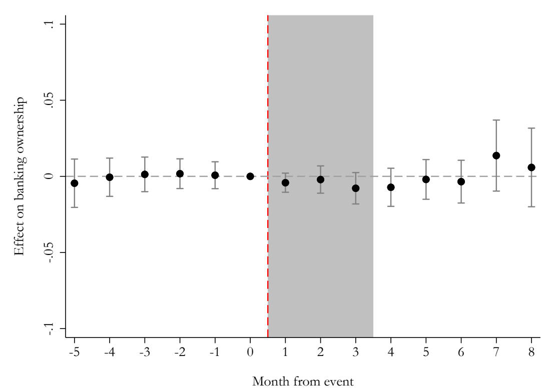
(b) Fintech loans



(c) Mobile phones ownership



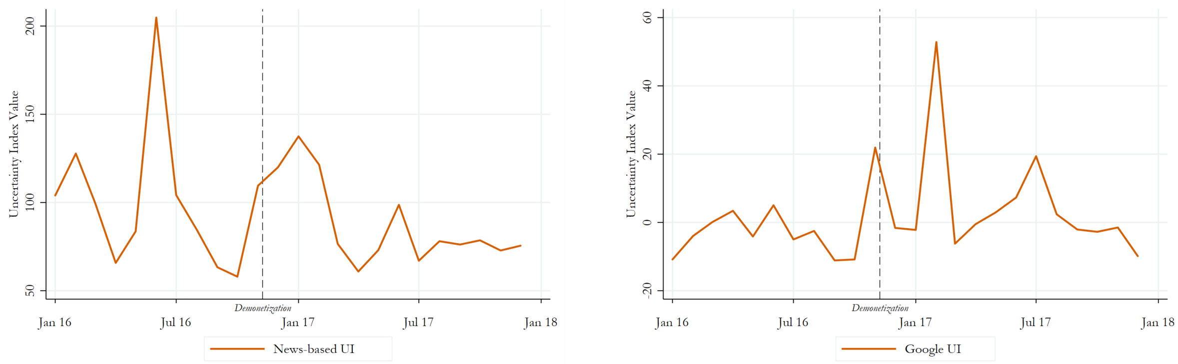
(d) Bank account ownership



Notes: The figure plots the dynamic effects on adoption of different technologies across districts based on distance of that district to the closest district with more than 500 active firms before the Demonetization. The main outcomes of interest are: log of the new firms joining the platform in that month (panel (a)) which essentially replicates our main finding; log of number of loans applied that month on a leading fintech company (panel (b)); the share of households with mobile phone in the district as reported in CMIE (panel (c)); the share of households with a bank account in the district as reported in CMIE (panel (d)). The approach is the same already followed in Figure 7: each figure reports the result from the dynamic difference-in-difference using the distance-to-hub treatment, where the month before the shock is normalized to zero (October 2016). We employ the dichotomous treatment, which is equal to one if a district is further than 400 *kms*. from the closest payment hub. 95% confidence intervals are represented with the vertical lines; standard errors are clustered at the district level.

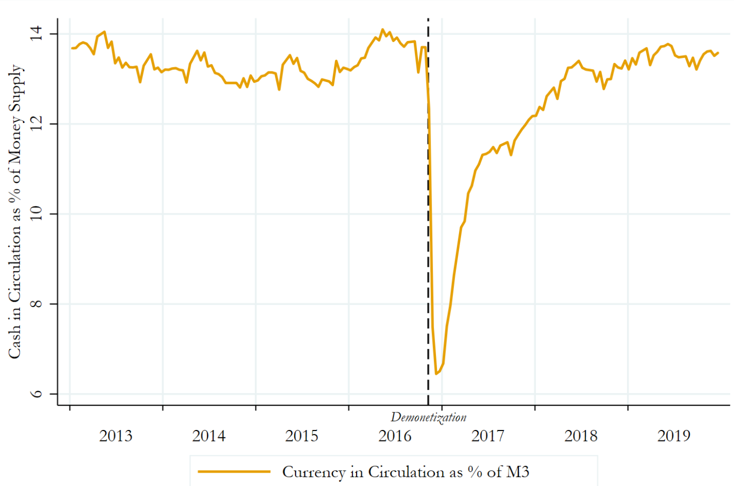
Figure H.11: Uncertainty indices for India (2016-2017)

(a) News-based uncertainty index (b) Google search-based uncertainty index



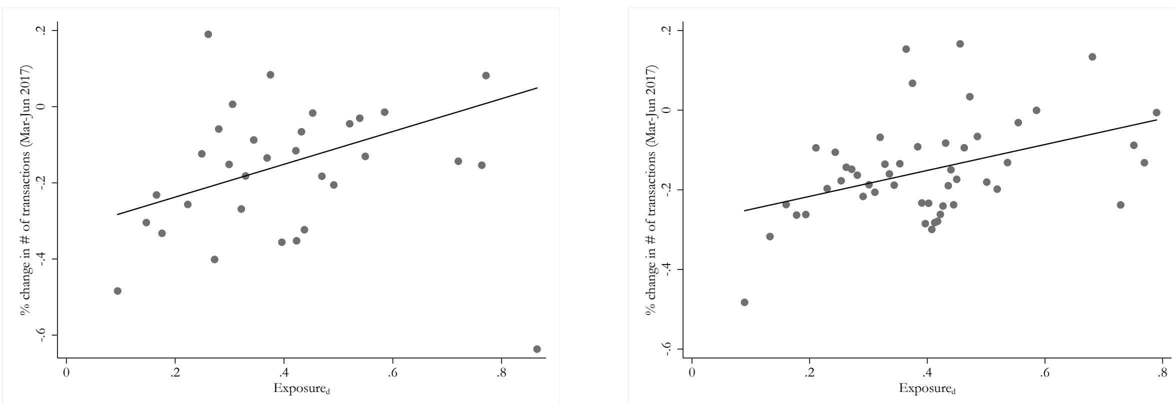
Notes: The figure reports the monthly uncertainty measures between January 2016 and December 2017 as constructed in Priyaranjan and Pratap (2020) . The vertical line is the month of the Demonetization (i.e. November 2016). Panel (a) reports the series constructed by extracting uncertainty sentiment from newspapers. Panel (b) reports the series constructed by extracting uncertainty sentiment from Google Search data. Detailed discussion provided in Section A.1 of the paper and in Priyaranjan and Pratap (2020).

Figure H.12: Cash in circulation



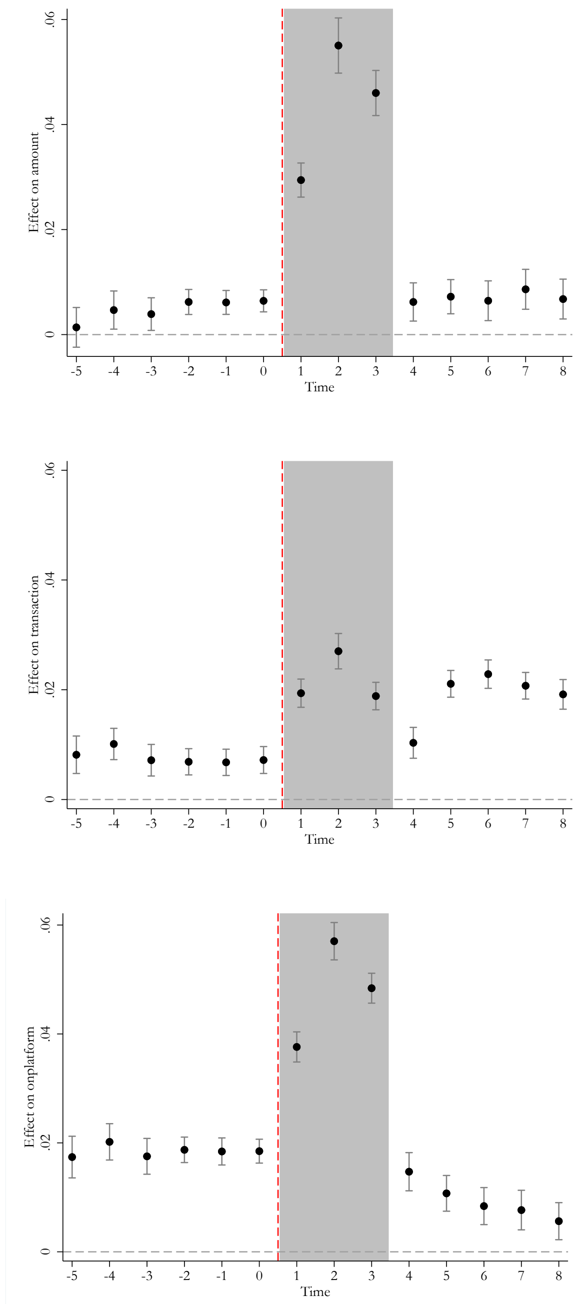
Notes: The figure reports the monthly ratio between cash in circulation and money supply (M3) between 2013 and 2020. The vertical line is in between October 2016 and November 2016. Source: Reserve Bank of India.

Figure H.13: Exposure$_d$ and early-adopters



Notes: The figure reports the results of the analysis on early-adopters, defined as those retailers that have adopted between November 2016 and January 2017, and are still active in March 2017. The figure plots the relationship between the growth in number of transactions between March 2017 and June 2017 at firm-level ΔY_{id} and the size of the shock Exposure$_d$. Specifically, we estimate equation $\Delta Y_{id} = \beta$ Exposure$_d + \Gamma X_d + \epsilon_d$, where ΔY_{id} is the growth rate in $Y_{id,t}$ between $t =$ March 2017 and $t =$ June 2017. $Y_{id,t}$ are the number of transactions in month t for firm i, located in district d. Panel (a) reports the baseline analysis, while panel (b) also controls for the level conducted during the Demonetization period.

Figure H.14: Firm adoption dynamics in electronic payments data based on existing adopters



Notes: The figure plots month-by-month estimates of the dependence of firm-level adoption rates on the share of other adopters in the industry/pincode. The specification we estimate is a version of equation 111 in which each coefficient is interacted with a weekly dummy; we reported the monthly estimates of the coefficient γ. The top panel reports the effects when x is the total amount of transactions, the middle panel reports the effects when x is the total number of transactions, and the bottom panel reports the effects when x is a dummy for whether the firm used the platform over the past week. 95% confidence intervals are represented with the vertical lines; standard errors are clustered at the pincode level.

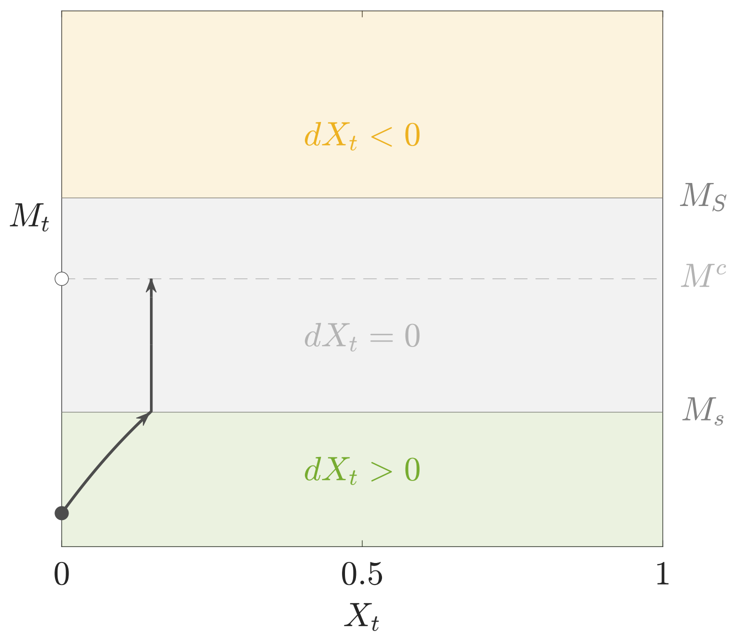


Figure H.15: Adoption dynamics in the fixed cost model of Appendix B.6. The figure reports the phase diagram of the model. The gray solid lines represent the adoption thresholds M_s and M_S, and the dashed grey line indicates the long-run level of cash demand, M_c. The green region corresponds to the states of the economy where firms currently using cash adopt electronic money at rate k ($dX_t = (1 - X_t)kdt$), while the yellow regions correspond to the state of the economy where firms currently using electronic money adopt cash at rate k ($dX_t = -X_t kdt$), as described in Result 2. The grey region is the inaction region. The solid arrow illustrates a potential trajectories of the economy following a large drop in cash demand, from $M_{0^-} = M_c$ (the hollow marker) to $M_0 \ll M^c$ (the solid marker). In this hypothetical trajectory, innovations to cash demand for $t > 0$ are exactly zero, so that $M_t = \mathbb{E}_0[M_t|M_0]$.

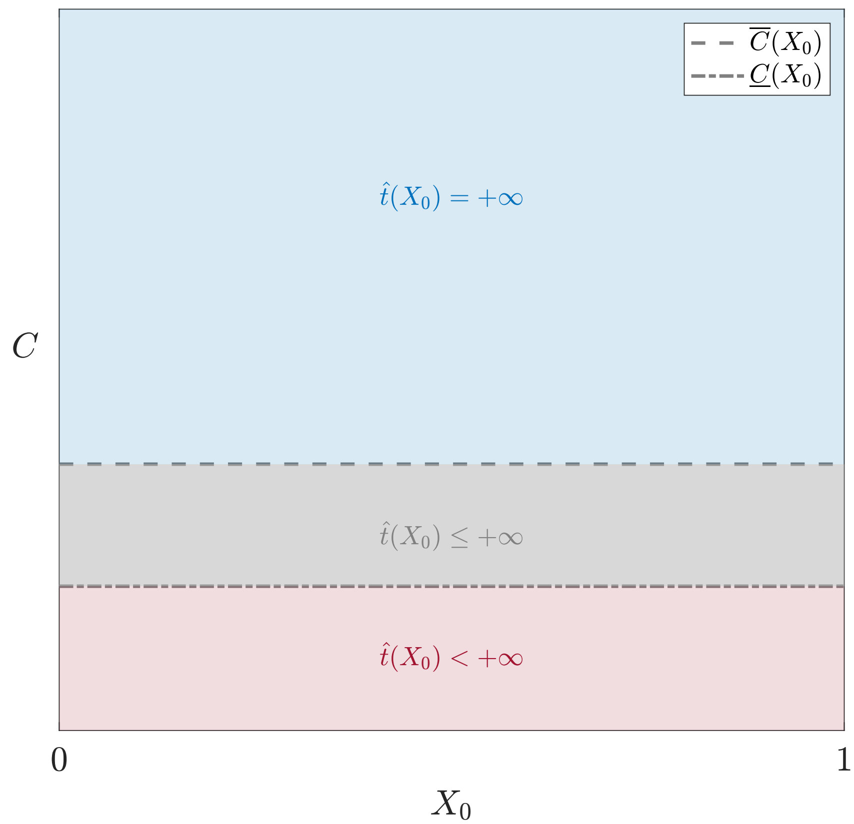
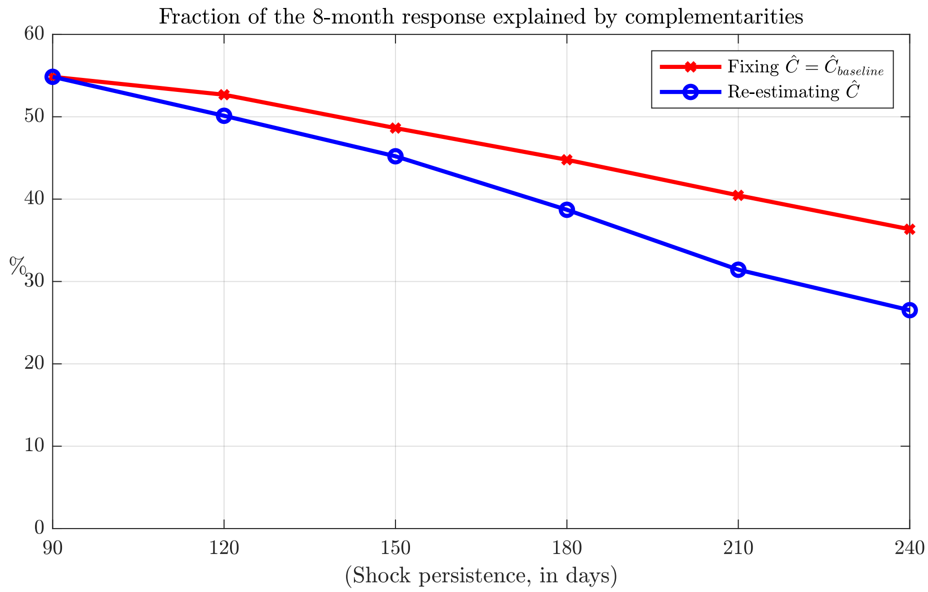


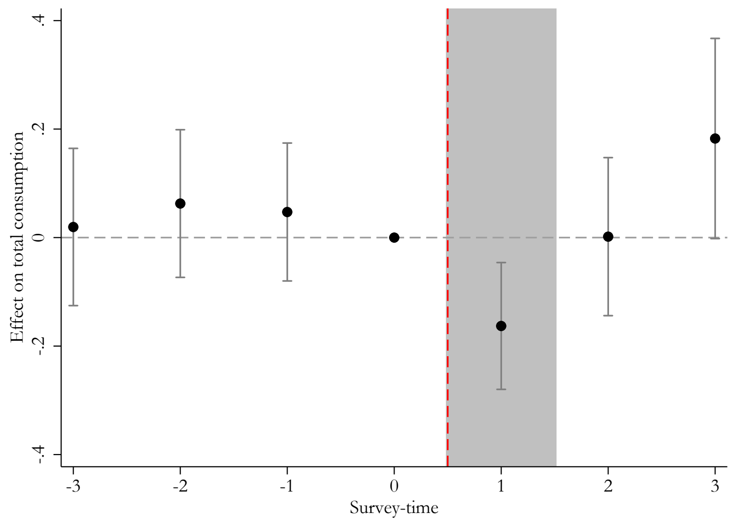
Figure H.16: Summary of perfect foresight response to a large shock when $\theta \leq k$. The shock size S is assumed to satisfy $S > M_c^{-1}(1 + k/(r + k))(M^c - M^e)$. The region highlighted in red corresponds to values of (X_0, C) such that adoption stops at a finite time in the perfect foresight response to a shock of size S. The blue area corresponds to values of (X_0, C) such that adoption continues at all dates $t \geq 0$. In the gray are, the bounds on the equilibrium adoption threshold are not sufficiently tight to determine whether adoption stops at a finite horizon or whether the shock leads to adoption at all future dates. The parameter values used to construct the graph are: $r = -\log(0.70)/12$ (the calibration is monthly); $k = 0.200$; $M_c = 1$; $M_e = 0.970$; $\theta = -30\log(1 - 0.90)/240$; $\sigma = 0.06$; $T = 1200$. In this calibration, $\theta \leq k$ (the shock is mean-reverting quickly). Figure 3 summarize the perfect foresight response when $\theta > k$. See Appendix B.2 for derivations of $\hat{t}(X_0)$ and the boundaries $\underline{C}(X_0)$ and $\overline{C}(X_0)$.

Figure H.17: Persistence of fundamentals and the role of complementarities



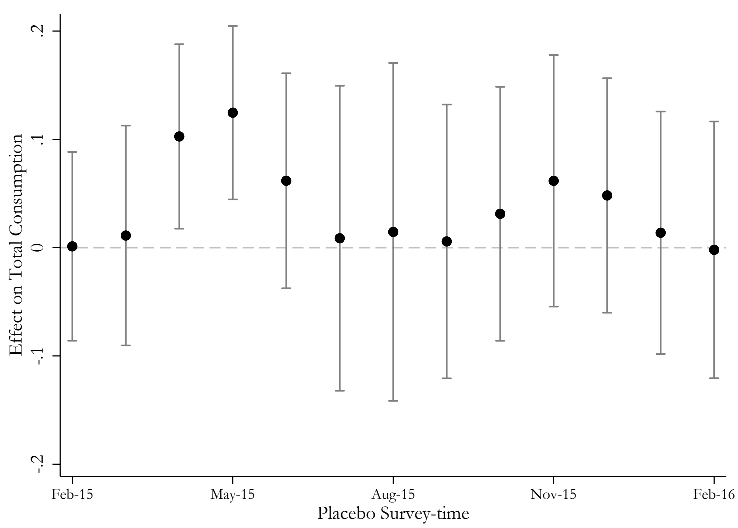
Notes: The figure shows the estimated contribution of complementarities to the 8-month response of adoption to the shock, under different assumptions about the persistence of changes in the flow benefits of cash. Persistence (expressed as the expected time for cash-based demand to converge back to within 10% of its long-run value) is on the horizontal axis. The contribution of complementarities (expressed as one minus the ratio of adoption response 8 months after the shock when $C = 0$, to the adoption response 8 months after the shock when $C = \hat{C}$, where $\hat{C}$ is an estimate of C) is on the vertical axis. The red line reports the contribution of complementarities when varying the degree of shock persistence but keeping the estimates of complementarities equal to the value estimated in Section 5. The blue line reports the contribution of complementarities when we re-estimate different values of the parameter C under alternative assumptions about the persistence of the shock.

Figure H.18: Consumption responses based on exposure to the shock



Notes: The figure plots estimates of consumption responses depending on exposure to the shock (Exposure$_d$). The specification we estimate is a version of equation 109 in which each coefficient is based on the interaction of the treatment variable with a event-time dummy. We report the event-time estimates of the coefficient δ. The treatment is our measure of Exposure$_d$ as described in Section 4. The dependent variable on the y-axis is the (log) total expense by household (as described in Section D). 95% confidence intervals are represented with the vertical lines; standard errors are clustered at the district level. Source: CMIE Consumption Data.

Figure H.19: Consumption responses based on placebo shocks



Notes: The figure plots the estimates of consumption responses depending on exposure to the shock where we assume the occurrence of a "fake" shock in each survey-time corresponding to each entry on the x-axis. The specification we estimate is a version of equation 109 in which each coefficient is based on the interaction of the treatment variable (Exposure$_d$) with an event-time dummy. We report the coefficient δ for the event-time right after shock. The treatment variable is our measure of Exposure$_d$ for the district (as described in Section 4). The dependent variable $\log(y_{h,d,t})$ is the log of total consumption (as described in Section D). 95% confidence intervals are represented with the vertical lines; standard errors are clustered at the district level. Source: CMIE Consumption Data.

Table H.1: Share of Chest Banks and Deposit Growth

	Δ log(deposits)		Δ log(deposits$^{adj.}$)		Δ log(deposits^{N})	
	(1)	(2)	(3)	(4)	(5)	(6)
Chest Exposure	0.093***	0.082***	0.085***	0.074***	1.573***	1.383***
	[0.013]	[0.012]	[0.013]	[0.012]	[0.222]	[0.206]
log(Pre Deposits)		-0.035***		-0.035***		-0.591***
		[0.003]		[0.003]		[0.055]
% villages with ATM		0.054		0.052		0.909
		[0.061]		[0.063]		[1.022]
% villages with banks		-0.064**		-0.065**		-1.085**
		[0.029]		[0.030]		[0.492]
Rural Pop./Total Pop.		-0.065***		-0.072***		-1.088***
		[0.016]		[0.018]		[0.277]
log(population)		0.037***		0.035***		0.616***
		[0.003]		[0.003]		[0.058]
Observations	512	512	512	512	512	512
R-squared	0.118	0.314	0.099	0.291	0.118	0.314
District Controls		✓		✓		✓

Notes: The table reports the results from regression of the district-level deposit growth (between September 30, 2016 and December 31, 2016) on the measure of Exposure$_d$ for the district (as described in Section 4). Columns (1) and (2) use the measure of change in total deposits. Columns (3) and (4) uses the measure of abnormal growth in total deposits, which adjust for the normal deposit growth in the district across the last two years. Specifically, we subtract the mean deposit growth in the last 8 quarters from the growth in 2016Q4 deposits. Columns (5) and (6) uses the dependent variable of deposit growth that is normalized to have mean zero and standard deviation 1. Odd columns shows the correlation without any controls. Even columns include the district-level controls for (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population and level of population in the district. Robust standard errors are reported in parentheses; $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.2: Exposure$_d$ and Deposit Growth (pre-shock quarters)

	(1) 201604	(2) 201603	(3) 201602	(4) 201601	(5) 201504	(6) 201503	(7) 201502	(8) 201501	(9) 201404	(10) 201403	(11) 201402	(12) 201401
Chest Exposure	1.383***	-0.303	0.263**	0.176	0.163	0.328	0.021	0.289	0.394	-0.653**	0.183	0.094
	[0.206]	[0.217]	[0.120]	[0.189]	[0.227]	[0.208]	[0.197]	[0.184]	[0.281]	[0.258]	[0.210]	[0.222]
Observations	512	512	512	512	512	512	512	512	512	512	512	512
R-squared	0.314	0.023	0.039	0.129	0.025	0.080	0.043	0.047	0.025	0.042	0.120	0.092
District Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression of district-level deposit growth for all eleven quarters before the shock (2016 Q4) on the density of chest banks in the district. The dependent variable is normalized to have mean zero and standard deviation 1. Treatment variable is our measure of Exposure$_d$ for the district (as described in Section 4). District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population and level of population in the district. Standard error in parentheses; $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.3: Relation between Exposure$_d$ and district's technology penetration

Dependent variable:	coeff. (1)	R^2 (2)
Amount per firm (in '1000 Rs.)	0.072 (0.052)	0.192
# transactions per firm	0.258 (0.228)	0.149
Share of active firms	0.006 (0.008)	0.261
Credit card ownership rate	0.020 (0.045)	0.055
Mobile phone ownership rate	0.047 (0.031)	0.062
Bank account ownership rate	0.031 (0.030)	0.057

Notes: The table reports the correlation between our treatment Exposure$_d$ and measures of ex-ante technology penetration in a district. Column (1) reports the conditional effect of our exposure measure on the variable listed in that specific row, using the same controls as the main analysis. In the first three rows, we report measures of penetration of our mobile wallet technology. We measure the amount of transactions (in thousand rupees), number of transactions, and number of retailers in the platform, all scaled by the total number of businesses with less than 4 employees in the district (2013 Economic Census). The last three rows examine the relationship for measures of penetration of other technologies that are relevant for the use of electronic payment: share of households that have credit card, mobile phone, and bank account (obtained from CMIE data). Column (2) reports R^2 and standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.4: Path dependence

$y =$	log(amount)	log(# users)	log(# switchers)	log(amount)	log(# users)	log(# switchers)
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure$_d$ $\times$ $\mathbf{1}_{t \geq t_0}$	1.481***	0.489***	0.401***	2.584***	0.871***	0.667***
	[0.411]	[0.130]	[0.122]	[0.699]	[0.261]	[0.230]
$L1.y$	0.603***	0.779***	0.635***			
	[0.015]	[0.008]	[0.008]			
$L2.y$				0.271***	0.494***	0.294***
				[0.021]	[0.015]	[0.010]
Observations	6,656	6,656	6,656	6,144	6,144	6,144
R-squared	0.906	0.954	0.899	0.860	0.912	0.848
District f.e.	✓	✓	✓	✓	✓	✓
Month f.e.	✓	✓	✓	✓	✓	✓
District Controls $\times$ Month f.e.	✓	✓	✓	✓	✓	✓

Notes: The table replicates the main results in the paper studying the exposure to the shocks and technology adoption (i.e. Table 3) also controlling for lagged outcomes. Columns (1)-(3) include a one-month lag, while Columns (4)-(6) include a two-months lag. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.5: Electronic Payment and Cash shock: non-linear effects

	log(amount)	log(# users)	log(# switchers)
	(1)	(2)	(3)
$\mathbf{1}(\text{Exposure Quintile} = 2)_d \times \mathbf{1}_{t \geq t_0}$	0.493	-0.036	-0.088
	[0.410]	[0.206]	[0.163]
$\mathbf{1}(\text{Exposure Quintile} = 3)_d \times \mathbf{1}_{t \geq t_0}$	0.734	0.176	0.090
	[0.451]	[0.221]	[0.170]
$\mathbf{1}(\text{Exposure Quintile} = 4)_d \times \mathbf{1}_{t \geq t_0}$	1.434***	0.547**	0.364**
	[0.449]	[0.233]	[0.179]
$\mathbf{1}(\text{Exposure Quintile} = 5)_d \times \mathbf{1}_{t \geq t_0}$	1.496***	0.503**	0.335*
	[0.476]	[0.231]	[0.176]
Observations	7,168	7,168	7,168
R-squared	0.851	0.869	0.819
District f.e.	✓	✓	✓
Month f.e.	✓	✓	✓
District Controls × Month f.e.	✓	✓	✓

Notes: The table examines the relationship between the exposure to the shock and the use of electronic payment, using our treatment variable non-parametrically. The baseline specification is the same as our main specification in the paper (i.e. Table 3). In all Columns we replace our continuous treatment with a set of dummies that divide the sample in five quintiles based on the size of the exposure measure. The group in the first quintile of exposure measure is excluded and acts as reference group. All columns include the district level controls interacted with month dummies, as well as district and month fixed-effects. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.6: District adoption rates based on initial adoption in electronic payment data

	log(amount)		log(# users)		log(# switchers)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}\ (\text{Any Adopter})_d \times \mathbf{1}\ (t \geq t_0)$	1.449***		1.698***		1.278***	
	[0.367]		[0.185]		[0.145]	
$\log(\text{pre-amount})_d \times \mathbf{1}\ (t \geq t_0)$		0.053		0.166***		0.122***
		[0.049]		[0.022]		[0.017]
Observations	7,168	7,168	7,168	7,168	7,168	7,168
R-squared	0.851	0.849	0.880	0.877	0.830	0.826
District f.e.	✓	✓	✓	✓	✓	✓
Month f.e.	✓	✓	✓	✓	✓	✓
District Controls × Month f.e.	✓	✓	✓	✓	✓	✓

Notes: The table shows adoption dependence on initial conditions at the district level. The specification estimated is equation 110. In the first row, I_d is a dummy if a district had a positive adoption level before the Demonetization. In the second row, I_d the total amount transacted before the Demonetization. Across the six columns, we focus on different measures of activity in the platform. Specifically, we examine: in Columns (1) and (2), the total amount (in Rs.) of transactions carried out using a digital wallet in district d during month t; in Columns (3) and (4), the total number of active retailers using a digital wallet in district d during month t; in Columns (5)-(6), the total number of new retailers joining the digital wallet in district d during month t. District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population and level of population in the district. Standard errors are clustered at the district level. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.7: Distance to hub_d and district characteristics (Balance Test)

Dependent variable:	coeff. (1)	R^2 (2)
Chest Exposure	-0.004	0.303
	(0.062)	
Log(Pre Deposits)	-0.587	0.268
	(0.378)	
% villages with ATM	0.019	0.709
	(0.014)	
# Bank Branches per 1000's	-0.010	0.540
	(0.009)	
# Agri Credit Societies per 1000's	0.000	0.306
	(0.019)	
% villages with banks	-0.004	0.895
	(0.014)	
Log(Population)	-0.306	0.460
	(0.193)	
Literacy rate	0.000	0.548
	(0.028)	
Credit card ownership rate	-0.075	0.394
	(0.054)	
Mobile phone ownership rate	-0.015	0.287
	(0.030)	
Bank account ownership rate	-0.045	0.238
	(0.044)	

Notes: The table tests for differences in observable district-characteristics and the distance-to-the-hub measure $(\text{Distance to hub})_d$. In Column (1). we report the coefficient from the OLS regression of each variable on the distance-to-the-hub measure, controlling for the distance to the state-capital and corresponding state fixed-effect. Standard errors clustered at the district level are reported in parentheses. We also report the R^2 of the analysis in the Column (2). $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.8: Heterogeneity of the response across distance from the hubs

	log(amount)	log(# users)	log(# switchers)
	(1)	(2)	(3)
$\text{Exposure}_d \times \mathbf{1}_{t \geq t_0}$	6.290**	2.152*	1.647
	[2.548]	[1.264]	[1.032]
$\mathbf{1}_{D_d \geq 100km} \times \mathbf{1}_{t \geq t_0}$	4.411**	1.907**	1.587**
	[1.886]	[0.930]	[0.774]
$\text{Exposure}_d \times \mathbf{1}_{D_d \geq 100km} \times \mathbf{1}_{t \geq t_0}$	-6.116**	-2.404*	-1.907*
	[2.574]	[1.286]	[1.054]
Observations	7,168	7,168	7,168
R-squared	0.887	0.911	0.871
District f.e.	✓	✓	✓
Month f.e.	✓	✓	✓
District Controls × Month f.e.	✓	✓	✓
State × Month f.e.	✓	✓	✓

Notes: The table analyzes the heterogeneous effects of the exposure to the shock (Exposure_d) across districts far and close to the electronic payment hubs. We define as districts closer to the hub those locations that are within 100 km from the closest hub. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.9: Abnormal technology growth around government policies

	November 8, 2016	December 8, 2016	January 1, 2017
	(1)	(2)	(3)
Exposure_d	0.376**	0.046	-0.151
	[0.187]	[0.173]	[0.139]
Observations	512	512	512
R-squared	0.169	0.031	0.034
District Controls	✓	✓	✓

Notes: The table uses district-level, high-frequency data to check for abnormal growth in adoption of digital wallet around the announcement or introduction of government policies. We estimate the equivalent of our difference-in-difference collapsed across periods: $\Delta y_d = \beta \text{Exposure}_d + \gamma X_d + \epsilon_d$ where Δy_d is the two-week symmetric growth in transaction following the date specified in the header. To provide a benchmark, in column (1) we examine the behavior around the Demonetization announcement date (essentially replicating our main finding using a different specification). In column (2), we instead examine the response around the date when the new government policies were announced and in columns (3) when the policies were implemented. The controls are the same as the main analysis. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.10: Distance results controlling for fintech familiarity

	log(amount)		log(# users)		log(# switchers)	
	(1)	(2)	(3)	(4)	(5)	(6)
(Distance to hub)$_d$ $\times$ $\mathbf{1}(t \geq t_0)$	-3.795***	-4.301***	-1.637***	-1.786***	-1.082***	-1.198***
	[1.144]	[1.139]	[0.481]	[0.487]	[0.372]	[0.376]
Observations	7,168	7,168	7,168	7,168	7,168	7,168
R-squared	0.887	0.889	0.912	0.913	0.871	0.873
District f.e.	✓	✓	✓	✓	✓	✓
Month f.e.	✓	✓	✓	✓	✓	✓
District Controls× Month f.e.	✓	✓	✓	✓	✓	✓
State× Month f.e.	✓	✓	✓	✓	✓	✓
# pre loans × Month f.e.		✓		✓		✓

Notes: The table replicates the main result examining the relationship between distance from the hub and electronic payment after including an extra control that captures the local propensity to adopt new fintech products, which we measure as number of fintech loans per capita in the district before November 2016. Odd columns replicates the finding already reported in the paper (Table 3). Even columns show how the result changes when we control for our proxy for the propensity to adopt technology interacted with month dummies. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.11: District adoption rates based on initial adoption: Alternative specification

	log(amount)			log(# users)			log(# switchers)		
	$\delta = 200$	$\delta = 300$	$\delta = 400$	$\delta = 200$	$\delta = 300$	$\delta = 400$	$\delta = 200$	$\delta = 300$	$\delta = 400$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Distance To Hub > δ km.) $\times$ $\mathbf{1}_{\{t \geq t_0\}}$	-1.291***	-1.123***	-1.104***	-0.528***	-0.489***	-0.484***	-0.346**	-0.351***	-0.353***
	[0.373]	[0.356]	[0.344]	[0.182]	[0.151]	[0.139]	[0.143]	[0.115]	[0.107]
Observations	7,168	7,168	7,168	7,168	7,168	7,168	7,168	7,168	7,168
R-squared	0.887	0.887	0.887	0.912	0.912	0.912	0.871	0.872	0.872
District f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Month f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
District Controls $\times$ Month f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓
State $\times$ Month f.e.	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table replicates Table 3 (i.e. use of the technology as a function to the distance to the nearest hub), where we proxy distance with a dummy variable, rather than a continuous one. In other words, the specification estimated is equation 16, replacing D_d with a dummy for distance to hub based on threshold δ ($\mathbf{1}_{\{\text{Distance To Hub}>\delta \text{ km.}\}}$). The different thresholds are specified at the top of the table (i.e. 200 km, 300 km, and 400 km). Apart from this alternative definition, the rest of the specification is the same as Table 3. Standard errors clustered at the district level are reported in parentheses. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.12: Firm adoption based on existing adoption rate in electronic payments data

	(1)	(2)	(3)	(4)
	$x_{i,k,p,t} = \log(\text{amount})_{i,k,p,t}$			
$x_{i,k,p,t-1}$	0.528	0.437	0.369	0.358
	(0.005)	(0.004)	(0.004)	(0.004)
$X_{k,p,t-1}$	0.090	0.155	0.032	0.015
	(0.003)	(0.001)	(0.001)	(0.001)
R^2	0.365	0.404	0.455	0.460
	$x_{i,k,p,t} = \log(\#\text{ transactions})_{i,k,p,t}$			
$x_{i,k,p,t-1}$	0.707	0.617	0.593	0.577
	(0.005)	(0.005)	(0.005)	(0.005)
$X_{k,p,t-1}$	0.032	0.062	0.041	0.017
	(0.002)	(0.002)	(0.001)	(0.001)
R^2	0.549	0.574	0.601	0.606
	$x_{i,k,p,t} = \mathbf{1}\{\text{On platform}\}_{i,k,p,t}$			
$x_{i,k,p,t-1}$	0.509	0.404	0.334	0.323
	(0.005)	(0.004)	(0.003)	(0.003)
$X_{k,p,t-1}$	0.046	0.097	0.038	0.022
	(0.004)	(0.003)	(0.002)	(0.001)
R^2	0.341	0.387	0.443	0.448
Firm F.E.		✓	✓	✓
Pincode × Week F.E.			✓	✓
Industry × Week F.E.				✓
Observations	11,750,558	11,750,558	11,541,757	11,541,757

Notes: The table reports estimates of the dynamic specification for adoption based on : $x_{i,p,k,t} = \alpha_i + \alpha_{p,t} + \alpha_{k,t} + \rho x_{i,p,k,t-1} + \gamma X_{p,k,t-1} + \epsilon_{i,p,k,t}$ allowing for spillovers across industries within the same pincode p (specification (111)). We reported estimates of the coefficient γ. The top panel reports effects when x is the total value of transactions, the middle panel reports effects when x is the total number of transactions, and the bottom panel reports effects when x is a dummy for whether the firm used the platform in the week. Standard errors clustered at pincode level are reported in parentheses. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.13: Firm adoption based on existing adoption rate (allowing for spillovers across industries)

	(1)	(2)	(3)	(4)
		$x_{i,p,d,t} = \log(\text{amount})_{i,p,d,t}$		
$x_{i,p,d,t-1}$	0.533	0.444	0.375	0.358
	(0.006)	(0.005)	(0.004)	(0.003)
$X_{p,d,t-1}$	0.076	0.135	0.023	0.016
	(0.002)	(0.002)	(0.002)	(0.001)
R^2	0.364	0.402	0.432	0.441
		$x_{i,p,d,t} = \log(\#\text{ transactions})_{i,p,d,t}$		
$x_{i,p,d,t-1}$	0.711	0.621	0.586	0.579
	(0.005)	(0.005)	(0.005)	(0.005)
$X_{p,d,t-1}$	0.022	0.043	0.021	0.013
	(0.001)	(0.001)	(0.001)	(0.001)
R^2	0.548	0.573	0.585	0.590
		$x_{i,p,d,t} = \mathbf{1}\{\text{On platform}\}_{i,p,d,t}$		
$x_{i,p,d,t-1}$	0.496	0.381	0.334	0.323
	(0.007)	(0.003)	(0.003)	(0.003)
$X_{p,d,t-1}$	0.035	0.071	0.027	0.015
	(0.002)	(0.001)	(0.001)	(0.001)
R^2	0.347	0.398	0.420	0.428
Firm F.E.		✓	✓	✓
Industry × Week F.E.			✓	✓
District × Week F.E.				✓
Observations	11,750,558	11,750,558	11,750,558	11,749,732

Notes: The table reports estimates of the dynamic specification for adoption based on : $x_{i,p,d,t} = \alpha_i + \alpha_{dt} + \rho x_{i,p,d,t-1} + \gamma X_{p,d,t-1} + \epsilon_{i,p,d,t}$ allowing for spillovers across industries within the same pincode. We reported estimates of the coefficient γ. The top panel reports effects when x is the total value of transactions, the middle panel reports effects when x is the total number of transactions, and the bottom panel reports effects when x is a dummy for whether the firm used the platform in the week. Standard errors are clustered at the pincode level. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.14: Firm adoption based on existing adoption rate (district-level)

	(1)	(2)	(3)	(4)
		$x_{i,k,d,t} = \log(\text{amount})_{i,k,d,t}$		
$x_{i,k,d,t-1}$	0.572***	0.474***	0.420***	0.410***
	(0.0100)	(0.0108)	(0.0108)	(0.0105)
$X_{k,d,t-1}$	0.0696***	0.117***	0.0295***	0.00606***
	(0.00257)	(0.00662)	(0.00439)	(0.00134)
R^2	0.398	0.437	0.459	0.463
		$x_{i,k,d,t} = \log(\#\text{ transactions})_{i,k,d,t}$		
$x_{i,k,d,t-1}$	0.776***	0.709***	0.635***	0.624***
	(0.0101)	(0.00933)	(0.0149)	(0.0148)
$X_{k,d,t-1}$	0.0237***	0.0600**	0.116***	0.0212***
	(0.00821)	(0.0301)	(0.00693)	(0.00205)
R^2	0.598	0.615	0.635	0.637
		$x_{i,k,d,t} = \mathbf{1}\{\text{On platform}\}_{i,k,d,t}$		
$x_{i,k,d,t-1}$	0.528***	0.408***	0.378***	0.370***
	(0.00828)	(0.00931)	(0.00857)	(0.00849)
$X_{k,d,t-1}$	0.0158***	0.0314***	0.0198***	0.00489***
	(0.00131)	(0.00180)	(0.00202)	(0.000938)
R^2	0.369	0.419	0.433	0.437
Firm F.E.		✓	✓	✓
District × Week F.E.			✓	✓
Industry × Week F.E.				✓
Observations	58,022,429	58,022,429	58,021,662	58,021,662

Notes: The table reports estimates of the dependence of firm-level adoption rates on the share of other adopters in the industry/district. The specification we estimate is a version of equation 111 at *district-level* in which each coefficient is interacted with a weekly dummy; we reported estimates of the coefficient γ. The top panel reports effects when x is the total value of transactions, the middle panel reports effects when x is the total number of transactions, and the bottom panel reports effects when x is a dummy for whether the firm used the platform in the week. Standard errors are clustered at the district level. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.15: Effect of exposure on pre-adopters

	log(amount)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Exposure}_d \times \mathbf{1}(t \geq t_0)$	1.370**		1.348**		1.418**	
	[0.598]		[0.633]		[0.699]	
$\text{Exposure}_d \times \mathbf{1}(\text{Short run})_t$		1.613**		1.625**		1.670**
		[0.634]		[0.675]		[0.711]
$\text{Exposure}_d \times \mathbf{1}(\text{Long run})_t$		1.224**		1.182*		1.268*
		[0.602]		[0.631]		[0.718]
$\text{Exposure}_d \times \mathbf{1}(\text{High SCI})_d \times \mathbf{1}(t \geq t_0)$			-0.883			
			[2.056]			
$\text{Exposure}_d \times \mathbf{1}(\text{High SCI})_d \times \mathbf{1}(\text{Short run})_t$				-1.225		
				[2.147]		
$\text{Exposure}_d \times \mathbf{1}(\text{High SCI})_d \times \mathbf{1}(\text{Long run})_t$				-0.678		
				[2.018]		
$\mathbf{1}(\text{High SCI})_d \times \mathbf{1}(t \geq t_0)$			0.903	0.903		
			[1.316]	[1.316]		
$\text{Exposure}_d \times \mathbf{1}(\text{High Lang. Conc.})_d \times \mathbf{1}(t \geq t_0)$					-0.246	
					[1.131]	
$\text{Exposure}_d \times \mathbf{1}(\text{High Lang. Conc.})_d \times \mathbf{1}(\text{Short run})_t$						-0.316
						[1.109]
$\text{Exposure}_d \times \mathbf{1}(\text{High Lang. Conc.})_d \times \mathbf{1}(\text{Long run})_t$						-0.204
						[1.157]
$\mathbf{1}(\text{High Lang. Conc.})_d \times \mathbf{1}(t \geq t_0)$					-0.113	-0.113
					[0.644]	[0.644]
Observations	132,608	132,608	132,552	132,552	132,454	132,454
R-squared	0.543	0.544	0.544	0.544	0.544	0.544
Firm f.e.	✓	✓	✓	✓	✓	✓
Month f.e.	✓	✓	✓	✓	✓	✓
District Controls × Month f.e.	✓	✓	✓	✓	✓	✓

Notes: The table replicates the main effect of the shock using firm-level data and focusing only on pre-adopters (i.e. firms that have adopted before the November 2016 Demonetization shock). The specification is the equation (15), with the exception that district-fixed effects are replaced by firm-fixed effects. The dependent variable is the log of amount transacted by the firm on the platform during month t. In Column (1), we replicate the main finding of the paper with this sample. In Column (2), we decompose the effect splitting the post-November effect between the short-run (i.e. November 2016 to January 2017) and long-run (i.e. February 2017 onwards). In Columns (3) and (5), we report the interaction between our treatment and the post-shock dummy with proxies of learning. In columns (4) and (6), we report the interaction between the short- and long-run effect with proxies of learning. We consider two proxies for learning: social connectivity of the district (Columns 3 and 4); and measure of language concentration (Columns 5 and 6). For both proxies, we split the sample at median to identify regions with high/low learning. Standard errors clustered at the district level are reported in parentheses. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.16: Effect of exposure on adoption by districts' language concentration and social connectedness

	log(# switchers)		
	(1)	(2)	(3)
Exposure$_d$ $\times$ $\mathbf{1}(t \geq t_0)$	0.768**	0.754**	0.751**
	[0.330]	[0.342]	[0.340]
Exposure$_d$ $\times$ (Lang. Conc.)$_d$ $\times$ $\mathbf{1}(t \geq t_0)$		0.224	
		[0.374]	
(Lang. Conc.)$_d$ $\times$ $\mathbf{1}(t \geq t_0)$		-0.161	
		[0.197]	
Exposure$_d$ $\times$ SCI$_d$ $\times$ $\mathbf{1}(t \geq t_0)$ SCI$_d$			-0.283
			[0.316]
SCI$_d$ $\times$ $\mathbf{1}(t \geq t_0)$			0.130
			[0.134]
Observations	6,874	6,874	6,874
R-squared	0.822	0.822	0.822
District f.e.	✓	✓	✓
Month f.e.	✓	✓	✓
District Controls $\times$ Month f.e.	✓	✓	✓

Notes: The table reports estimates of the effect of cash contraction on the adoption of digital wallet, after controlling for and interacting with proxies of learning in a district. The outcome considered is the number of firms joining the platform that month. In Column (1), we replicate the main finding of the paper (equation 15). In Columns (2) and (3), we test whether there is any difference in this effect between districts where it is easier to learn. Specifically, in column (2) we use a proxy based on language concentration, while in column (3) we use a proxy based on social connectivity of the district on Facebook. More details on the analysis can be found in Section 4.5 and Appendix F. Standard errors clustered at the district level are reported in parentheses. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.17: Survey: Reasons for the Adoption of e-Payment Method Post-Demonetization

Reason for Adoption	Non-Business Owner	Small Business Owner	Total
New cash was hard to find	137	103	240
I learnt from my family or friends who have already used them	99	93	192
Shops where I buy things my customers started using non-cash payments	188	136	324
Any Reason	246	186	432

Notes: The table reports the key result from the survey analysis, discussed in the paper. The table shows the reasons reported by respondents for the reason that drove their decision to adopt a form of electronic payment in the aftermath of the Demonetization. For each sample considered, we report the total number of individuals in that group (i.e. any reason), as well as the number of respondents that select each of the three reasons: (1) new cash was hard to find; (2) I have learnt from my family or friends who have already use the form of electronic payment (i.e. learning); (3) Shops where I buy things (or "my customers" for business owner) started using non-cash payments (network effects). Notice that the options are not mutually exclusive and respondents can choose more than one. Results are reported separately for respondents that were (i) only customers of shops ("Non-Business Owner"), (ii) were also shopkeepers or business owners ("Small Business Owner"), and (iii) the combined sample ("Total"). More details on the survey are provided in Section F of the paper.

Table H.18: Population Distribution Comparison: Survey and 2011 Census of India

	Census 2011	Survey
Panel A: Age Comparison (%)		
Between 18 and 30 years old	34.3	36.4
Between 30 and 50 years old	40.5	60.1
Between 50 and 70 years old	20.0	3.3
70 years old or above	5.2	0.2
Panel B: By Region (%)		
Central Zone	23.4	7.0
Eastern Zone	21.5	3.9
North Eastern Zone	3.7	1.7
Northern Zone	13.1	9.0
Southern Zone	23.0	67.8
Western Zone	15.3	10.7

Notes: The table compares the geographical and age distribution from the survey and 2011 Census of India. Central Zone includes the states of Chhattisgarh, Madhya Pradesh, Uttar Pradesh and Uttarakhand. Eastern Zone includes the states of Bihar, Jharkhand, Odisha, and West Bengal. North Eastern Zone includes the states of Arunachal Pradesh, Manipur, Assam, Meghalaya, Mizoram, Nagaland, Tripura, and Sikkim. Northern Zone includes the states of Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir, Punjab, and Rajasthan. Southern Zone includes the states of Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu. Western Zone includes the states of Goa, Gujarat, and Maharashtra.

Frankel and Burdzy (2005)	This paper
Mode 1	Electronic money e
Mode 2	Cash c
Switching cost functions	$c^m(k^m, X) = 0, \quad m = 1, 2$
Payoff shocks	$W_t = -M_t$
Flow payoff in mode 1	$u(1, W_t, X_t) = M^e + CX_t$
Flow payoff in mode 2	$u(2, W_t, X_t) = -W_t$
Shock process	$\begin{cases} dW_t &= (\nu_t W_t + \mu_t)dt + \sigma_t dZ_t \\ \nu_t &= \begin{cases} \theta & \text{if} \quad t \leq T \\ 0 & \text{if} \quad t > T \end{cases} \\ \mu_t &= -\nu_t M^c \\ \sigma_t &= \sigma \end{cases}$
Relative flow payoff in mode 1	$D(W_t, X_t, k^1, k^2) = M^e + W_t + CX_t$
Lipschitz constants for D	$\beta = C, \quad \overline{\alpha} = 1$
Bounds on switching rates	$\underline{K}_1 = \underline{K}_2 = 0, \quad \overline{K}_1 = \overline{K}_2 = k$
Bounds on shocks process	$\begin{cases} N_1 &= \sigma \\ N_2 &= \frac{3}{2} \max(\sigma, \theta, M^c\theta, T\theta) \end{cases}$
Bound for strict dominance of mode 1	$\overline{w} = \left(\frac{r + k + \theta}{r + k + \theta e^{-(r+k+\theta)T}} \right) (M^c - M^e) - M^c$
Bound for strict dominance of mode 2	$\underline{w} = -\left(M^e + \frac{r + k + \theta}{r + k + \theta e^{-(r+k+\theta)T}} C \right)$

Table H.19: Mapping between the model of Section 3 and the general framework of Frankel and Burdzy (2005).

Table H.20: Consumption responses based on exposure to the shock

Exposure$_d$:	log(Expense$_{\text{Total}}$)	
	Continuous measure	Top 25%
	(1)	(2)
$(\text{Exposure})_d \times \mathbf{1}(t = t_1)$	-0.199***	-0.0577**
	(0.0637)	(0.0234)
$(\text{Exposure})_d \times \mathbf{1}(t = t_2)$	-0.0337	-0.0199
	(0.0815)	(0.0296)
$(\text{Exposure})_d \times \mathbf{1}(t = t_3)$	0.148	0.0146
	(0.102)	(0.0370)
$(\text{Exposure})_d \times \mathbf{1}(t = t_4)$	0.0252	-0.0187
	(0.141)	(0.0588)
Observations	564,690	564,690
R-squared	0.707	0.706
Household f.e.	✓	✓
Survey-time f.e.	✓	✓
District Controls × Survey-time f.e.	✓	✓
Household controls × Survey-time f.e.	✓	✓

Notes: The table shows the difference-in-differences estimate for consumption responses for each event-time after the demonetization shock relative to the pre-period (four event-time). The specification estimated is equation 109. The treatment variable is our measure of Exposure$_d$ for the district (Column (1)) and takes the values of 1 if the measure of Exposure$_d$ is in the top quartile of the distribution (Column (2)). The dependent variable $\log(y_{h,d,t})$ is the log of total consumption as defined in Section D. District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with a banking facility, share of rural population and level of population in the district. Household-level controls include pre-shock income and age of head of the household. Standard errors are clustered at the district level. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.21: Consumption responses across categories based on exposure to the shock

	Necessary	Unnecessary	Bills and Rent	Food	Recreation
	(1)	(2)	(3)	(4)	(5)
$(\text{Exposure})_d \times \mathbf{1}(t = t_1)$	-0.173***	-0.211**	0.253	-0.184***	-0.996**
	(0.0574)	(0.0989)	(0.268)	(0.0596)	(0.432)
Observations	564,690	564,690	564,690	564,690	564,690
R-squared	0.731	0.622	0.700	0.684	0.460
Household f.e.	✓	✓	✓	✓	✓
Survey-time f.e.	✓	✓	✓	✓	✓
District Controls × Survey-time f.e.	✓	✓	✓	✓	✓
Household controls × Survey-time f.e.	✓	✓	✓	✓	✓

Notes: The table shows the difference-in-differences estimate for consumption responses across various categories for each event-time after the demonetization shock relative the pre-period (four event-time). The specification estimated is equation 109. The treatment variable is our measure of Exposure$_d$ for the district (as described in Section 4). The dependent variable $\log(y_{h,d,t})$ is either the log of consumption of necessary goods (Column (1)); the log of consumption of unnecessary goods (Column (2)); log of expenditure on bills and rent (Column (3)); the log of expenditure on food (Column (4)); the log of expenditure on recreation activities (Column (5)) as defined in Section D. District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with a banking facility, share of rural population and level of population in the district. Household-level controls include pre-shock income and age of head of the household. Standard errors are clustered at the district level. $*** : p < 0.01$, $** : p < 0.05$, $* : p < 0.1$.

Table H.22: Consumption responses based on alternative cutoff for exposure to the shock

	log(Expense)		
	Total	Necessary	Unnecessary
	(1)	(2)	(3)
$\mathbf{1}_{\{t=t_1\}} \times$ (Top 25% Exposure)$_d$	-0.0577**	-0.0427*	-0.0781**
	(0.0234)	(0.0230)	(0.0343)
$\mathbf{1}_{\{t=t_2\}} \times$ (Top 25% Exposure)$_d$	-0.0199	-0.0172	-0.0277
	(0.0296)	(0.0266)	(0.0454)
$\mathbf{1}_{\{t=t_3\}} \times$ (Top 25% Exposure)$_d$	0.0146	-0.00438	0.0519
	(0.0370)	(0.0307)	(0.0533)
$\mathbf{1}_{\{t=t_4\}} \times$ (Top 25% Exposure)$_d$	-0.0187	-0.0588	0.0374
	(0.0588)	(0.0580)	(0.0786)
Observations	564,690	564,690	564,690
R-squared	0.706	0.731	0.622
Household f.e.	✓	✓	✓
Survey-time f.e.	✓	✓	✓
District Controls × Survey-time f.e.	✓	✓	✓
Household controls × Survey-time f.e.	✓	✓	✓

Notes: The table shows difference-in-differences estimate for consumption responses for each event-time post the demonetization shock relative the pre-period (four event-time). The specification estimated is equation 109. Treatment variable takes the value of 1 if our measure of Exposure$_d$ for the district (as described in Section 4) is in the top 25% value of exposure. The dependent variable $\log(y_{h,d,t})$ is either log of total consumption (Column (1)); log of consumption of necessary goods (Column (2)); log of consumption of unnecessary goods (Column (3)) as defined in Section D. District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population and level of population in the district. Household-level controls include pre-shock income and age of head of the household. Standard errors are clustered at the district level. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.

Table H.23: Heterogeneous consumption responses by district's exposure to alternate payment system

	Total		Necessary		Unnecessary	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}(t=t_1) \times (\text{Exposure})_d$	-0.303***		-0.298***		-0.280**	
	(0.0771)		(0.0740)		(0.121)	
$\mathbf{1}(t=t_2) \times (\text{Exposure})_d$	-0.177*		-0.201**		-0.114	
	(0.0972)		(0.0889)		(0.157)	
$\mathbf{1}(t=t_3) \times (\text{Exposure})_d$	0.103		0.0199		0.275	
	(0.131)		(0.108)		(0.203)	
$\mathbf{1}(t=t_4) \times (\text{Exposure})_d$	0.121		-0.124		0.445	
	(0.212)		(0.182)		(0.319)	
$\mathbf{1}(t=t_1) \times \mathbf{1}(\text{ATM})_d$	-0.118*	-0.0506*	-0.127**	-0.0491**	-0.0917	-0.0413
	(0.0615)	(0.0298)	(0.0518)	(0.0241)	(0.0963)	(0.0463)
$\mathbf{1}(t=t_2) \times \mathbf{1}(\text{ATM})_d$	-0.148**	-0.0535*	-0.157***	-0.0528**	-0.116	-0.0438
	(0.0663)	(0.0302)	(0.0584)	(0.0253)	(0.105)	(0.0490)
$\mathbf{1}(t=t_3) \times \mathbf{1}(\text{ATM})_d$	-0.0431	0.0000	-0.0481	-0.0105	-0.0118	0.0356
	(0.0731)	(0.0315)	(0.0615)	(0.0282)	(0.118)	(0.0486)
$\mathbf{1}(t=t_4) \times \mathbf{1}(\text{ATM})_d$	0.117	0.0644	-0.0285	-0.00480	0.299*	0.145*
	(0.114)	(0.0604)	(0.0968)	(0.0516)	(0.174)	(0.0864)
$\mathbf{1}(t=t_1) \times (\text{Top 25\% Exposure})_d$		-0.106***		-0.104***		-0.102**
		(0.0318)		(0.0291)		(0.0510)
$\mathbf{1}(t=t_2) \times (\text{Top 25\% Exposure})_d$		-0.0782**		-0.0829**		-0.0666
		(0.0375)		(0.0335)		(0.0609)
$\mathbf{1}(t=t_3) \times (\text{Top 25\% Exposure})_d$		0.00985		-0.0126		0.0669
		(0.0478)		(0.0385)		(0.0761)
$\mathbf{1}(t=t_4) \times (\text{Top 25\% Exposure})_d$		0.0227		-0.0993		0.211
		(0.0914)		(0.0776)		(0.137)
$\mathbf{1}(t=t_1) \times (\text{Exposure})_d \times \mathbf{1}(\text{ATM})_d$	0.185*		0.217**		0.129	
	(0.102)		(0.0987)		(0.151)	
$\mathbf{1}(t=t_2) \times (\text{Exposure})_d \times \mathbf{1}(\text{ATM})_d$	0.232*		0.261**		0.167	
	(0.119)		(0.112)		(0.185)	
$\mathbf{1}(t=t_3) \times (\text{Exposure})_d \times \mathbf{1}(\text{ATM})_d$	0.0744		0.0724		0.0536	
	(0.142)		(0.116)		(0.226)	
$\mathbf{1}(t=t_4) \times (\text{Exposure})_d \times \mathbf{1}(\text{ATM})_d$	-0.139		0.0818		-0.454	
	(0.218)		(0.202)		(0.326)	
$\mathbf{1}(t=t_1) \times (\text{Top 25\% Exposure})_d \times \mathbf{1}(\text{ATM})_d$		0.0913**		0.114***		0.0479
		(0.0437)		(0.0423)		(0.0645)
$\mathbf{1}(t=t_2) \times (\text{Top 25\% Exposure})_d \times \mathbf{1}(\text{ATM})_d$		0.101*		0.116**		0.0628
		(0.0520)		(0.0470)		(0.0811)
$\mathbf{1}(t=t_3) \times (\text{Top 25\% Exposure})_d \times \mathbf{1}(\text{ATM})_d$		0.00388		0.0110		-0.0363
		(0.0600)		(0.0494)		(0.0933)
$\mathbf{1}(t=t_4) \times (\text{Top 25\% Exposure})_d \times \mathbf{1}(\text{ATM})_d$		-0.0622		0.0771		-0.289*
		(0.0994)		(0.0949)		(0.148)
Observations	554,894	554,894	554,899	554,899	554,899	554,899
R-squared	0.704	0.704	0.730	0.730	0.618	0.618

Notes: The table shows triple-difference estimate for consumption responses for each event-time post the demonetization shock relative the pre-period (four event-time), based on district's access to ATM facility. Treatment variable is our measure of Exposure_d for the district (odd columns) and takes the values of 1 if the measure of Exposure_d is in the top quartile of the distribution (even columns). $\mathbf{1}(\text{ATM})_d$ takes the values of 1 if the number of ATM per capita in district is above the median of the distribution. The dependent variable $\log(y_{h,d,t})$ is either the log of total consumption (Column 1-2); log of necessary consumption (Column 3-4); log of unnecessary consumption (Column 5-6). as defined in Section D. District-level controls include (log) pre-shock banking deposits, share of villages with ATM facilities, share of villages with banking facility, share of rural population and level of population in the district. Household-level controls include pre-shock income and age of head of the household. Standard errors are clustered at the district level. $***: p < 0.01$, $**: p < 0.05$, $*: p < 0.1$.