One Shock, Many Disruptions: Firm Experience After India's Demonetization

Manisha Goel[1], Alexandra Schofield[2], Michelle Zemel[1]

[1] Pomona College Department of Economics, Claremont, CA, USA
[2] Harvey Mudd College Department of Computer Science, Claremont, CA, USA
manisha.goel@pomona.edu, xanda@cs.hmc.edu, michelle.zemel@pomona.edu

Abstract

We examine how firms describe their experiences in the aftermath of a sudden declaration in 2016 of demonetization in India that rendered 86% of cash in circulation no longer legal tender. We gauge firm exposure to the policy shock by the relative frequency of demonetization mentions in its financial reports. We also apply topic modeling to these reports to discern the different ways that firms were impacted. We find that firms are differentially exposed, with construction being most impacted and education and health services the least. Small firms are more exposed than large firms, although firms of all sizes and industries express uncertainty and worry about the future. Remarkably, even more than concerns about cash absence, the largest impact was uncertainty about the future.

Introduction

At 11:00 PM on November 8, 2016, India's Prime Minister announced in a nationally televised address that ₹500 and ₹1000 currency bills would be demonetized effective an hour later, at midnight. The removal of 86% of cash in circulation was an unanticipated shock for a cash-based economy: as of 2016, 90% of transactions in India were conducted in cash (Dharmapala and Khanna 2019). Estimates indicate that GDP declined 2% in the fourth quarter of 2016.

This aggregate macroeconomic shock could affect firms in the economy differently depending on the intensity and nature of their cash dependence. For example, the lack of cash could impede demand from customers and firms' ability to pay their employees and suppliers. Depending on a range of firm features, reliance on cash for certain categories of transaction could also change the effect of this policy on a firm. In this paper, we empirically examine the heterogeneous levels of exposures that Indian firms had to the demonetization policy using their own narrative accounts. We further examine these accounts to uncover the myriad pathways through which firms were affected.

Our research strategy relies on creatively combining structured financial data with unstructured textual data contained in firms' annual reports, applying natural language processing methods for two purposes. First, we determine how frequently a firm discusses demonetization in its narratives to quantitatively measure firm exposure to the policy shock. We document the heterogeneity in firm exposure, showing that exposure varies across both firm size and industry. Second, we use latent Dirichlet allocation (LDA) topic modeling (Blei, Ng, and Jordan 2003) to explore the myriad channels through which the policy disrupted firms. We develop a 50-topic model on text passages that discuss demonetization to uncover the ways firms reported disruptions to their operations and future plans. We uncover wide ranging disruptions, such as heightened uncertainty and pessimism about the macroeconomic outlook, absence of cash, falling product demand, and changing prices. We also quantify the extent to which each channel affects a given firm by measuring how much that firm discusses each topic from our model. We then document the heterogeneity in these topics, demonstrating how the channels of disruption vary across firms by industry and size.

Prior Work

In recent years, economic research has extensively focused on heterogeneous treatment effects. In a randomized or natural experiment, classical econometric approaches to estimate causal effects of treatment yield a single average treatment effect estimate (Imbens and Wooldridge 2009). However, the same treatment may have different effects on different sub-groups in the treatment group. Our paper shows that demonetization was another such natural experiment that affected firms to varying extents, as revealed by the heterogeneity in firm exposure. New methods have recently been developed to estimate heterogeneous treatment effects; see, for example, Wager and Athey (2018). However, once equipped with these estimates, researchers still have to rely on their intuition for why agents in their analyses are impacted differently. Our paper addresses this limitation by letting the data themselves speak: through distributions over co-occurring words, our estimated topics give us the reasons why firms are differently exposed. This approach also highlights the importance of using narrative data in economics, and social science research more broadly to tell us the reasons underlying these heterogeneous effects.

A small but growing literature examines the effect of the demonetization on Indian firms. Chodorow-Reich et al. (2020) show that economic activity, employment and bank credit fell significantly in the months after the policy announcement, and that businesses in the informal sector suffered severely. Dharmapala and Khanna (2019) examine how the stock market reacted to the announcement. Kisat and Phan (2020) show that consumer-facing firms suffered significantly as consumer demand dropped, but that this disruption did not spill over to upstream firms. Subramaniam (2020) shows that firms were unable to pay their employees and suppliers in the wake of demonetization.

We make three contributions to this literature. First, instead of selecting specific supply or demand challenges to investigate, we use the data to inform us directly about the many pathways through which demonetization disrupted firm operations and plans. Second, we examine how exposure and pathways of disruption differ across firm size and industries. Third, to the best of our knowledge, we are the first to use direct text analysis of firms' narratives as captured in their reports to investigate demonetization's consequences for India's economy.

Data, Measurement, and Identification

Our analysis combines structured data from the Prowess database,[1] which provides annual financial information for Indian firms, with firms' annual reports retrieved from the National Stock Exchange (NSE)[2] and Bombay Stock Exchange (BSE)[3] websites. We use optical character recognition (OCR) to extract raw text from annual report pdfs. Using annual firm financial and narrative data, we build a (unbalanced) panel dataset of 16,622 firm-years, comprised of 4,857 unique firms followed during 2016-2019.

Each firm-year observation includes firm size as measured by total assets and firm industry based on its National Industry Classification (NIC). The original NIC codes are aggregated into ten broad industry classifications: agriculture; arts and recreation; education and healthcare services; finance, insurance, and real estate; information and communication; manufacturing; mining, construction, and utilities; professional, technical, and administrative services; transportation and accommodation; and wholesale and retail trade.

Exposure analysis

To measure firm exposure to the policy, we calculate the proportion of word tokens in an annual report consisting of variants of "demonetization" (e.g., demonetisation, demonetizing, etc.). This approach, now widely adopted in economics, was proposed by Baker, Bloom, and Davis (2016). Since annual reports are long ($>$100 pages), the proportion of "demonetization" words is tiny. For ease of interpretation, we multiply these proportions by a million. Thus, for example, an exposure of 0.00001 is reported as 10.

Topical analysis of demonetization disruption

To understand the content of firms' demonetization-related experiences, we identify all passages in firms' annual reports that reference the policy. A demonetization passage is defined as all words within a 100-word radius of a key demonetization term. If two or more such terms exist within 100 words of each other, the boundary of the passage is extended until no key demonetization term appears within 100 words. We analyze the collection of all demonetization passages to uncover the different ways in which firms experience the shock. Passages span 3,453 firm-year observations from 2,069 unique firms observed during 2016-2019.

To analyze this collection of text passages we use latent Dirichlet allocation to model our text as a mixture of *topics*, or probability distributions over a vocabulary of terms. LDA is an unsupervised procedure and, thus, does not require document labels to learn a model; instead, it is parameterized only by the number of topics to infer, K, and the discrete vocabulary of words V for the text collection over which frequencies will be measured in each document. Empirically, inferred topics often correspond to human-interpretable subjects that can be labeled meaningfully by domain experts, which enables us to identify common themes (topics) in demonetization-related discussions. To infer our model, we use Mallet (McCallum 2002) to implement a collapsed Gibbs sampling inference procedure for LDA (Griffiths and Steyvers 2004). While neural NLP language models such as transformers might produce more accurate predictions of the language in these reports, their complex architecture and high volume of parameters make it difficult to find what phenomena cause particular model associations. We choose LDA as a more interpretable way to decompose the language of these reports, as the trained model is parameterized directly through probabilities over a fixed vocabulary across text documents.

A significant challenge in using computational methods with large text collections such as ours is pre-processing, or pruning a text collection to remove elements that would interfere with meaningful analysis. We determined a preprocessing workflow through iteration to determine if topics clearly focused around distinctive report content. Before training the topic model, we removed capitalization and punctuation, deleted duplicate texts, combined frequent multi-word phrases into single vocabulary items, and removed words that appeared overly frequently as well as extremely rarely. We chose not to apply a lemmatizer or stemmer to remove word affixes in order to avoid possibly losing meaningful signal from varying word morphology (Schofield and Mimno 2016).

We developed a 50-topic model trained with automatic optimization of asymmetric hyperparameters (Wallach, Mimno, and McCallum 2009). Two authors individually labeled each topic based on a combination of a list of the top 50 words and a sample of passages from the report that were dominantly of that topic. The two annotators then met and compared their labels to reach a consensus on the final topic labels.

The 50 topics were further classified into thirteen topic groups representing distinct themes based on consensus around which probabilistic topics represented similar subject areas. Twelve of these topic groups represent channels through which the policy shock affected firms. [4] These chan-

[1] https://prowessiq.cmie.com/

[2] https://www.nseindia.com/

[3] https://www.bseindia.com/

[4] In contrast to models that automatically learn topic hierarchies Li and McCallum (2006), our manual approach helps us to retain

nels are as follows: cash transactions, credit, consumer demand, fintech, current macroeconomic environment, capital markets, future macroeconomic outlook and uncertainty, prices of inputs and products, real estate, savings portfolios, supply chain, and workers. The final topic group aggregates topics whose themes do not describe channels of disruption (see for example, topics 10 and 40 in Tables 2-3).

In addition to surfacing the channels of disruption as reported by firms, the topic model also produces a per-report vector containing the proportion of the demonetization passages allocated to each of the 50 topics. This firm-year level quantification allows us to determine which firms shared common causes for disruption at which times. A full list of the 50 topics, their key words, and their topic group classification is provided in the Appendix in Tables 2-3).

Results

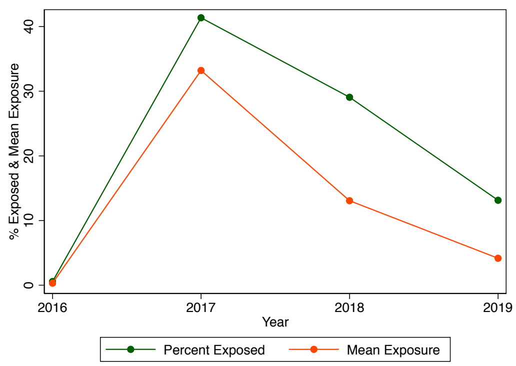

Figure 1: Percent firms exposed and their mean exposure. Firm exposure is defined as the proportion of report terms that are variants of the term "demonetization". This percentage is multiplied by 10^6 for ease of interpretation. The percent exposed measure calculates the percentage of firms that report non-zero exposure per year. The mean exposure averages all firms with non-zero exposure in a year.

Policy Shock Exposure Across Time and Firms

Using the relative frequency of demonetization terms in a firm's annual financial report as a measure of its exposure to the policy shock, we can observe how this shock was experienced over time and across firms. As Figure 1 shows, a small subset of firms mention demonetization related terms beginning in 2016. This small subset is constituted by firms that filed their annual reports in November or December 2016. Most firms file their annual reports for 2016-17 financial year in March 2017, making 2017 the first time we observe firm reports after the policy shock. The figure shows that in 2017, over 40% of firms mentioned demonetization in their reports. By 2019, this number fell to about 12%. The figure also shows the average firm exposure level, conditional on mentioning demonetization at all. This average exposure peaks in 2017 and then steadily decreases through 2019.

contextual differences between topics addressing similar themes.

Quartile	Exposure
1	96.72
2	89.30
3	79.69
4	72.09

(a) Exposure by size

Industry	Exposure
agriculture	57.03
mining, construction and utilities	86.88
manufacturing	68.84
wholesale and retail trade	77.94
transportation and accommodation	61.30
information and communications	82.26
finance, insurance, and real estate	107.12
professional, technical, and administrative services	87.07
education and health services	55.90
arts and recreation	74.28

(b) Exposure by industry

Table 1: Mean exposure by size quartile and industry. Firms are classified into size quartiles per year & then by industry. Exposure for each firm is defined as the proportion of words in the report that are demonetization terms. Exposure is averaged across firms with non-zero exposure in each category (quartile or industry), then multiplied by 10^6 for ease of interpretation.

To see how exposure differs across firms of different sizes, we divide firms into four size (assets) quartiles in each year and then calculate mean exposures across firms in each quartile in a given year. Results in Table 1(a) for the year 2017 show that exposure falls almost linearly across size quartiles, with the smallest firms being the most exposure.

Next, we examine how exposure differs across industries. Figure 1(b) shows that the highest exposure levels are felt by firms in finance, insurance, and real estate, followed by those in mining, construction, and utilities. This is consistent with our intuition. In India, a large proportion of transactions in real estate construction and associated services occur in cash. Thus, the absence of cash severely disrupted these sectors. Further, since banks and other financial institutions were having to deal with the influx of demonetized cash, and were witnessing an increased demand for credit, these firms were also significantly exposed to the policy shock. Agriculture and education and health industries were much less exposed in comparison.

Mechanisms Underlying Exposure

We see from the above discussion that there is considerable heterogeneity in demonetization exposure across firm size and industry. We now investigate *how* firms were exposed, i.e., the channels of exposure. We infer these channels from the topics and their proportions trained on the passages around demonetization-related terms in firms' annual financial reports, as described in the previous section.

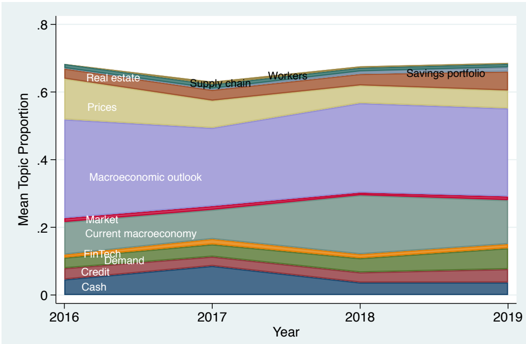

Figure 2: Mean Topic Proportions. The average proportion of a firm's demonetization passages attributable to each of the twelve topic groups is shown for 2016-2019. Proportions are determined through a weighted average approach using the summed topic proportions for each topic group and passage weighted by length of the corresponding passage.

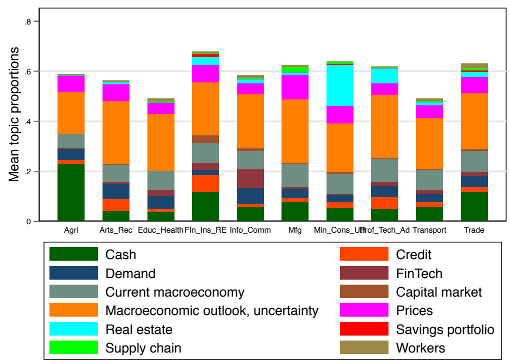

Figure 4: Mean topic proportions by industry. The proportion of a firm's demonetization passages attributable to each of the twelve topic groups is averaged by industry. The proportion allocated to each topic is averaged over all firms with non-zero exposure in 2017 in the same industry.

Figure 2 presents the mean proportions across all firms of the twelve topic groups created out of the original fifty topics. The means are presented for the period 2016-19. We have two key takeaways. First, topic proportions stay quite stable over the years, indicating that the nature of disruptions caused by the policy shock is persistent. Second, the largest disruptions are heightened uncertainty about the future along with a bleak economic outlook. There is also considerable concern about the current state of the economy. Changes in prices, disruptions created directly by absence of cash, and lack of consumer demand follow in their proportions.

Next, we examine how these disruptions compare across firm size. As before, we divide firms into four asset quartiles and calculate across firms in each quartile the mean proportions of topic groups. Our findings for the year 2017 are presented in Figure 3. The figure shows that across all quartiles, firms similarly expressed their concerns about the current state of the macroeconomy and uncertainty about the

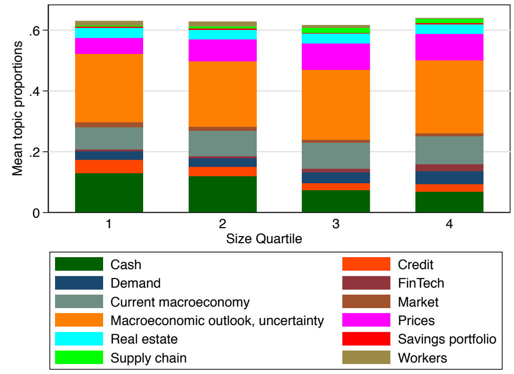

Figure 3: Mean topic proportions by firm size quartile. The proportion of a firm's demonetization passages attributable to each of the twelve topic groups is averaged by firm size. Firms are assigned to size quartiles each year based on total assets. The proportion allocated to each topic is averaged over all firms with non-zero exposure in the same size quartile during the year 2017.

future. Larger firms are somewhat less concerned about the direct cash disruption than smaller firms. This is consistent with our intuition since larger firms may have access to more sophisticated transaction mechanisms making them less reliant on cash. Similarly, they are also less concerned about credit than smaller firms. However, they express more concern about capital markets and goods and services prices.

Figure 4 presents mean topic proportions by industry for the year 2017. Firms in all industries express concerns about the current and future macroeconomy to comparable extents. Agriculture is most severely impacted by the absence of cash, followed by wholesale & retail trade and finance, insurance, and real estate (as explained earlier, real estate is particularly cash reliant in India). Mining, construction, and utility firms express considerable real estate related disruptions. Manufacturing firms are particularly concerned about input and output prices. Information and communication firms express the maximum disruption created by capital market volatility and credit disruption is expressed most by finance, insurance, and real estate firms.

Conclusion

By combining structured and unstructured data for Indian firms, we show that the single demonetization shock caused heterogeneous impacts for firms – both in the level and pathways of disruption. A key finding is that even more than the direction disruption created by the absence of cash, firms were impacted by the poor macroeconomic conditions the policy caused as well as heightened uncertainty about the future economic state. Smaller firms were more exposed than larger firms, but larger firms did not remain unscathed by any means. Our next steps include breaking down the sentiment associated with demonetization via both topics and a financial report-appropriate sentiment lexicon (Loughran and McDonald 2011).

Acknowledgments

We would like to thank Eloise Burtis, Matthew Ivler, Chris Nardi, and Asya Sklyar for their help on computing for this project and Pomona College for financial support.

References

Baker, S. R.; Bloom, N.; and Davis, S. J. 2016. Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4): 1593–1636.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3(null): 993–1022.

Chodorow-Reich, G.; Gopinath, G.; Mishra, P.; and Narayanan, A. 2020. Cash and the economy: Evidence from India's demonetization. *The Quarterly Journal of Economics*, 135(1): 57–103.

Dharmapala, D.; and Khanna, V. S. 2019. Stock Market Reactions to India's 2016 Demonetization. *Journal of Empirical Legal Studies*, 16(2): 281–317.

Griffiths, T. L.; and Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National aAademy of Sciences*, 101(suppl 1): 5228–5235.

Imbens, G. W.; and Wooldridge, J. M. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1): 5–86.

Kisat, F.; and Phan, M. 2020. Consumer Demand Shocks & Firm Linkages: Evidence from Demonetization in India. *Working Paper (SSRN 3698258).*

Li, W.; and McCallum, A. 2006. Pachinko allocation: DAG-structured mixture models of topic correlations. In *Proceedings of the 23rd International Conference on Machine learning*, 577–584.

Loughran, T.; and McDonald, B. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1): 35–65.

McCallum, A. K. 2002. Mallet: A MAchine Learning for LanguagE Toolkit. Http://mallet.cs.umass.edu.

Schofield, A.; and Mimno, D. 2016. Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics*, 4: 287–300.

Subramaniam, G. 2020. The Supply-Side Effects of India's Demonetization. *Working Paper (SSRN 3472758).*

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Wallach, H. M.; Mimno, D. M.; and McCallum, A. 2009. Rethinking LDA: Why priors matter. In *Advances in neural information processing systems*, 1973–1981.

Appendix

	α	Label	Topic Group	Topic Keys
0	0.0347	non-bank credit	credit	financial credit nbfcs sector banks
1	0.18629	firm future challenges and opportunities	outlook and uncertainty	business company growth focus year
2	0.07566	government provided infrastructure	not relevant	government infrastructure sector india development
3	0.11117	cash transactions	cash	cash year business impact demonetization
4	0.01194	macro outlook	outlook and uncertainty	economy growth india's fiscal global
5	0.00823	India's eight core infrastructure industries	not relevant	oil growth gas economy indian
6	0.02658	bank loans, microfinance	credit	loan loans portfolio bank credit
7	0.02676	steel	not relevant	steel demand cement industry production
8	0.05641	audit statement	not relevant	report financial statements company analysis
9	0.02516	capital market volatility	market	markets year market global indian
10	0.01532	agriculture	not relevant	farmers agriculture farm milk crop
11	0.01837	shift away from physical assets to financial assets	savings portfolio	financial savings insurance assets funds
12	0.02212	media and entertainment	not relevant	growth industry advertising media entertainment
13	0.01222	audit statement	not relevant	company management report discussion analysis
14	0.15132	macro impact - growth slowdown	macro current	growth economy india economic gst
15	0.04015	bankruptcy reform, npa	not relevant	bankruptcy sector code banks insolvency
16	0.02261	policy description - GST and demonetization	not relevant	tax gst growth governance compliance
17	0.00952	toll collection, transportation	not relevant	toll lakhs project collection march
18	0.03562	audit/compliance	not relevant	report management discussion analysis director
19	0.05527	macro outlook	outlook and uncertainty	growth gdp cent quarter year
20	0.01354	compensation	workers	company remuneration managerial personnel relationship
21	0.07764	input prices	prices	company year prices due demand
22	0.07146	demonetization negative effect on firm performance	not relevant	year company tax previous profit
23	0.12644	government reforms	not relevant	tax gst economy india goods
24	0.14393	macro impact - growth slowdown	macro current	year growth gst financial half

Table 2: Full LDA topic list with labels (part 1). The topic group assignment is shown as well as the top 5 highest probability words in the topic for each of the 50 topics. One group of topics is categorized as not-relevant because they do not represent mechanisms of disruption.

	α	Label	Topic Group	Topic Keys
25	0.11989	consumer demand impact	demand	growth industry expected india market
26	0.01903	good discard	not relevant	pradesh states maharashtra uttar state
27	0.01305	alcoholic beverage and pharmaceutical industry	not relevant	pharmaceutical industry market products healthcare
28	0.07638	global macro outlook	outlook and uncertainty	economy growth india economic indian
29	0.02918	automobile industry outlook	outlook and uncertainty	industry growth vehicles sales segment
30	0.03185	financing reactions to cash crunch	credit	company limited capital shares equity
31	0.02022	bank notes, cash balance, company financials	cash	cash notes bank permitted hand
32	0.04035	fintech, digital wallet, banking apps	fintech	digital bank banking payments transactions
33	0.0279	textile industry	not relevant	textile industry cotton exports domestic
34	0.01107	gold, jewelery gems	not relevant	gold jewellery sugar india demand
35	0.07251	firm financial performance (income , balance sheet, etc.)	not relevant	year crore crores growth revenue
36	0.03092	supply chain disruption	supply chain	products business segment company market
37	0.06561	real estate, home ownership	real estate	real estate sector housing rera
38	0.07902	industry outlook	outlook and uncertainty	company business market opportunities demonetization
39	0.09915	global macro outlook	outlook and uncertainty	growth global economies economic economy
40	0.0349	risk and internal controls	not relevant	risk management internal control company
41	0.01189	tourism industry - government policies to boost	not relevant	tourism india travel hospitality hotels
42	0.02942	bank deposit flow, increased liquidity	cash	bank rate banks deposits rbi
43	0.09111	price impact of demonetization	prices	inflation growth fiscal prices deficit
44	0.02152	policy and implementation description, demonetisation	not relevant	notes money currency government black
45	0.00971	macro impact, real estate	real estate	sector improved confidence real consumer
46	0.03117	firm strategy, beauty products retail	not relevant	products retail brand sales consumer
47	0.01931	employee training, csr	not relevant	bank demonetization branches training employees
48	0.03935	general statement of firm performance to shareholders	not relevant	shareholders year annual report company
49	0.01538	macro outlook	outlook and uncertainty	india trillion cent expected billion

Table 3: Full LDA topic list with labels (part 2).