Essays on Industrial Organization of Banking and Finance

Vinod Dharmarajan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Dong-Jae Eun, Chair

Philip Brock

Quan Wen

Program Authorized to Offer Degree:

Economics

University of Washington

Abstract

Essays on Industrial Organization of Banking and Finance

Vinod Dharmarajan

Chair of the Supervisory Committee:
Dong-Jae Eun, Chair
Department of Economics

I explore the market structure of the Indian banking industry and its implications for designing policies. My main chapter (Chapter 2) analyses how multi-market contact between banks (measured as overlap between banks across different marketplaces) affects the market structure. I find evidence of an oligopolistic market, with a dominant bank.

In Chapter 1, I measure the competition in the Indian banking industry using the Lerner index and Boone indicator. I construct a unique dataset of base rates of banks at a quarterly cadence to estimate different measures of competition in the market. Past studies have used imputed price of loans, calculated as the ratio of interest earned over total amount of loans and advances. Such an approach has its limitations when banks experience large amount of non-performing loans and advances and leads to lower imputed prices even when the actual price of loans has not decreased. I address this shortcoming by using the dataset of bank-wise quarterly base rates. I find competition is decreasing in the banking industry during

the sample period whereas using imputed price of loans shows increasing competition.

In Chapter 2, I empirically test for evidence for the theory of mutual forbearance or the linked oligopoly theory. The theory states that firms compete less aggressively with each other in a given market if they fear retaliation from rivals in other markets where they both operate. There are two main results in this paper: (1) I look at whether multimarket contact between banks in India due to geographic overlap of branches affects bank profitability. Using a bank-level fixed effects model for a sample of 110 unique banks between 2005-2021, I find evidence supporting the theory of mutual forbearance. (2) I use a Bresnahan-Lau structural model framework to estimate the market structure parameter for the Indian banking industry. Defining markets at the state-level and focusing on the market for credit for the same time period, I estimate these structural parameters using a simultaneous, non-linear 3SLS system of equations with instrumental variables. The estimated structural parameters indicate that the market represents one where prices (interest rates for loans) are set under joint profit maximization. I argue that rather than a monopoly market, the banking market is characterized by a dominant bank who sets the market price and the remaining banks who set prices based off the residual demand curve. This is based on the data on market share, where the largest bank has a market share of at least 16% throughout the sample period (irrespective of whether market share is measured in terms of deposits, credit, assets or number of branches).

Chapter 3, explores the different determinants of access to formal finance as well as its usage at the household and individual level in India. Using a nationally representative survey data covering 2018, I find significant differences (economically and statistically) in terms of gender in access to and use of formal finance. At an individual level, I find that

women are 29% less likely to have a bank account relative to men. Further, at the level of the household, households headed by females are 9% less likely to have accessed credit from a formal source relative to households headed by men. I apply Cragg's Double Hurdle model and find evidence of the gender disparity in the intensity of use of credit from formal sources. Households headed by women use 29.9% less credit from formal sources and 10.6% credit from informal sources compared to households headed by men. Existence of gender disparities in access and use of finance call for greater attention while devising policies to increase financial inclusion and avoid accentuating these disparities.

Findings from the first two chapters indicate that the policy to expand branches could be counter-productive in improving financial access. Expansion of branches can increase multi-market contact, which reduces the competition in the market and can lead to higher interest rates on credit. In the third chapter, I find differences in access to and use of formal financial resources by gender, caste (social group), religion, education, and place of residence. Increasing branch networks can have the unintended effect of worsening financial access to minorities. Together, the findings call for greater attention on designing policies to address these structural issues on financial inclusion.

Table of Contents

List Of Figures

List Of Tables

Acknowledgements

I am incredibly thankful to my advisor, Professor Dong-Jae Eun. This dissertation would not be possible without his comments, advice, support, and encouragement. He has helped me find solutions when I had hit a dead end. I am also extremely grateful to my reading committee members, Professor Philip Brock and Professor Quan Wen, for their support and guidance throughout my PhD. Their intriguing questions has helped me improve and refine my research presented here. My research skills and my career as an economist have benefited tremendously from the numerous conversations with my advisors. I want to also thank the Graduate School Representative, Professor Prashanth Rajivan for taking interest in my work and providing critical feedback.

I am grateful to my wife, Abhiprerna, for her love and support. Her support and patience was indispensable in completing this dissertation. I would also like to thank my family - my parents, C.H. Dharmarajan and Sali Alex, as well as my brothers, Vivek and Abhiraj, for their encouragement and love. I am thankful to Dr. Samita Sareen for encouraging me to pursue this path towards a doctorate. I am also immensely thankful to my friends - Atul, Benjamin, Sleeba, Goutham, Madhav, Manasi, Ziqi, Kwong-yu, Agraj and Pushpak, for their support since I started this journey.

I am also thankful to Niketa and Tetyana in mentoring me while I was a Research Assistant at Landesa. Lastly, I am grateful to the Department of Economics at University of Washington and at Seattle University for giving me the opportunity to teach and hone my skills as an instructor.

Dedication

To my family

Chapter 1

Measuring Competition in the Indian Banking Industry

The banking industry in India offers a rich environment to study the impact of various policy changes on competition, efficiency and productivity. There has been quite a few policy changes since 1991 to increase competition in the Indian banking industry. These changes were introduced with the objective of improving the efficiency of banks in the industry, increasing access to finance and improve financial inclusion. In 1992, the monetary authority in India, the Reserve Bank of India (RBI) introduced changes in policy deregulating the market, particularly in terms of setting interest rates and allowing entry of private banks. In 1998, RBI introduced further reforms with the objective of strengthening financial stability. Another major change in policy occurred in 2005, where RBI made changes to how new branch licenses were issued. This change was made to increase competition between banks especially in the under-banked regions of the country.

According to Structure-Conduct-Performance school of thought, collusion is easier in concentrated markets [1]. Therefore, easing entry and requirements for opening new branches is expected to reduce the market concentration and in turn, improve competition between banks. Even if market concentration does not decrease, easing entry and exit restrictions is expected to increase contestability in markets. However, expansion of bank's branch network also provides banks with a new channel to collude among themselves. Moreover, mergers, acquisitions and amalgamations of banks has created a constantly evolving market structure.

[1] Bernheim and Whinston (1990), Feinberg (1984), Spagnolo (1999), Thomas and Willig (2006) provide theoretical framework supporting this hypothesis.

All previous studies (such as Prasad and Ghosh (2007) and Perera et al. (2006))have used the imputed price of loans (defined as the ratio of interest earned on loans and advances over the total amount of loans and advances) in their measures of competition. However, during this particular period, public sector banks suffered from increasing amounts of non-performing loans. As a result, computing the imputed price of loans can be misleading as the denominator, total amount of loans and advances, becomes reduces at a faster rate than the numerator when banks write off their non-performing loans. This incorrectly leads to a decrease imputed price of loans even though these banks might still be charging a higher interest rate for their loans. I address this issue by constructing a unique dataset of base rates set by every bank in the sample each quarter. Base rates are the interest rates on loans charged by banks to clients with zero credit risk. Lerner indices constructed using the base rates show that competition increased at the beginning of the sample period but has decreased in the banking industry since then.

Given continued consolidation of the public sector banks by the Government of India[2], the findings in this study call for greater attention to the competitive effects of such consolidation on the market. Decreasing competition can lead to banks keeping interest rates higher in under-banked market and reduce financial inclusion.

1.1 Background and Summary of the literature

Why study the Indian banking industry? The banking industry in India has a long history of dramatic policy changes over the past 70 years. Today, the Indian banking industry is characterized by a strong presence of public sector banks where the Government of India is the majority stakeholder. This is the result of decades of regulatory policy pursued since

[2]As of 01-Apr-2020, the number of public sector banks reduced from 27 to 12 (`https://www.livemint.com/industry/banking/merger-of-10-public-sector-banks-to-come-into-effect-from-today-10-points-1158563.html`).

independence. Starting in 1967, the government introduced policies such as the introduction of social controls to ensure a better alignment of the banking system to the needs of the nation's economic policy. In 1969, 14 scheduled commercial banks were "nationalized", i.e. taken over by the Government with the objective of "to better serve the needs of development of the economy in conformity with national policy objectives". In 1980, another six private sector banks were nationalized by the government. Between 1970 and early 1990's the Government and RBI imposed a wide range of price and quantity regulations. This created a scenario where large proportion of deposits were held in the form of reserves, and a strictly administered interest rate regime created a high cost and low quality of financial intermediation services.

However, in 1992, the government and RBI instituted financial reforms following the recommendations of the Narasimhan Committee report of 1991 to increase competition in the banking industry. The government allowed for greater entry of privately owned banks as well as foreign banks. Further, in 1998, a second set of reforms were introduced following recommendations of Narasimhan Committee report of 1998. This led to liberalizing the interest rate regime. The government's ownership in public sector banks were reduced and these banks were allowed to raise upto 49% of their capital from the equity markets. These reforms also rationalized various policy rates such as the statutory liquidity ratio, cash reserve ratio, and introduced an array of micro-prudential measures[3]. These reforms intended to create a level playing field for banks with different ownership[4]. Changes were also made to prudential norms related to capital adequacy, recognition of income, asset classification, norms for provisioning and exposures. There were two new private banks that entered the

[3]In this phase, distinctions in how regulations were applied by ownership were removed. Regulations on interest rates, reserve requirements and prudential rooms were made same for all banks. Priority Sector Lending remained the only category where the regulations and conditions were set differently for domestic and foreign banks. Domestic banks were required to allocated 40% if their lending to PSL while foreign banks were required to set 32% of their lending towards PSL. Note that the definition of priority sector was defined by RBI and changes over time.

[4]However, the requirements to lend to priority sectors were still different for domestic banks compared to foreign banks.

market in 2003 and 2004 respectively. There were further policy changes in 2005 and in subsequent years.

I focus on scheduled banks and more specifically on certain types of commercial scheduled banks. The Reserve Bank of India is the monetary authority in India and have regulatory authority over all scheduled banks. Scheduled banks are those banks that meet the criteria in section 42(6)(a) of the Reserve Bank of India Act, 1934. The advantage of the scheduled bank status is access to funding from the central bank at the bank rate as well as membership of the clearing house. Simply put, scheduled banks have access to more liquidity than non-scheduled banks as they can borrow short-term funds from central bank as well as participate in primary auctions of government securities undertaken by the central bank. Scheduled banks also become eligible to partner in government-run financial inclusion schemes.

Scheduled banks consists of commercial banks and cooperative banks. The commercial banks include 5 types - Public sector banks, Private sector banks, Foreign banks, Regional Rural banks, Small Finance Banks, and Payments banks. Until 2014, there were 4 categories of scheduled banks in India - Public Sector banks, Private Sector banks, Foreign banks, Regional Rural banks. The public sector banks are also referred to as nationalized banks as the Government of India is a majority stakeholder in these banks. Foreign Banks and Private Banks differ in terms of extent of foreign ownership with the latter having less than 49% in foreign ownership. Both public sector banks and private sector banks have to follow RBI's guidelines on allocation of credit to priority sectors as well as on branch expansion. Regional Rural banks (RRBs) are banks focused on rural credit. These banks are jointly owned by the Government of India, the state government (based on location of the bank) and a sponsor bank. RRBs were created as specialized rural financial institutions, catering predominantly to credit requirements of small borrowers in rural areas of the country.

Cooperative banks are different from commercial banks as they are registered as cooperative societies. The co-operative credit system can be divided into two sets of entities. The

first set focuses on ensuring flow of credit to the agricultural sector. It consists Primary Agricultural Credit Societies (PACS) at the village level, Central Cooperative Banks (CCBs) at the district level, and State Cooperative Banks (StCBs) at the state level[5]. PACS are not regulated by RBI. CCBs and StCBs are regulated by RBI along with another agency, NABARD (National Bank for Agricultural and Rural Development). The second set consists of Primary Cooperative Banks or Urban Cooperative Banks (UCBs) which are focused on financial requirements of clients in urban and semi-urban areas. In general, the regulation of cooperative banks by RBI was restricted to its functions directly related to banking activities. However, this has now changed with an amendment in 2020 to the Banking Regulations Act, 1949.

In 2014, RBI invited applications for universal banking licenses in the private sector as well introduced two new types of banking entities - Small Finance banks and Payments banks. Table 1.1 summarizes the key differences between Universal Banks (which includes all Public Sector Banks and Private Banks), Small Finance Banks, and Payments Banks.

The Indian banking industry is also unique as it is one largest social banking experiment in the world, where a substantial part of the industry is dominated by public sector banks, where the government is the majority stakeholder. Moreover, the Indian banking industry emerged relatively unscathed during the global financial crisis. This is partly because of the regulatory measures imposed by the RBI prior to and during the crisis. However, it is also interesting to study how banks coped with the crisis to help inform policy decisions in the future.

The existing literature on the Indian banking industry can be separated into two main topics: (a) studies measuring efficiency of Indian banks and changes in efficiency over time; (b) studies measuring competition in the banking industry and its changes overtime.

[5]In terms of administrative units, village is the lowest unit, followed by district and then the state.

Table 1.1: Characteristics of Universal Banks, Payments Banks and Small Finance Banks

	Universal Banks	Payments Banks	Small Finance Banks
Eligibility	Companies in the private and public sectors and non-banking financial companies (NBFCs) will be eligible to set up a bank through a wholly-owned non-operative financial holding company (NOFHC). These applicants need to meet the criteria set by RBI. The players will also need to have a sound and successful track record of 10 years.	Prepaid payment Instrument issuers, individuals/professionals, NBFCs, corporate business correspondents, mobile telephone companies, super-market chains, real sector cooperatives that are owned and controlled by residents, and public sector entities are eligible to apply for payments bank licenses. The promoter should be able to meet the 'fit and proper' criteria with a sound track record of of five years.	Resident individuals/professionals with 10 years of experience in banking and finance, companies and societies owned and controlled by residents, existing NBFCs, microfinance institutions, and local area banks can apply for small finance bank licenses. All entities should be owned and controlled by Indian residents and should be able to meet the 'fit and proper' criteria stated by RBI
Capital Requirement	The initial minimum paid-up voting equity capital for a bank needs to be at least INR 5 billion. The bank will need to be listed within three years of starting business.	The minimum paid-up equity capital for payments banks shall be INR 1 billion.	The minimum paid-up equity capital required is INR 1 billion.
Scope of Activity	The bank can accept deposits and carry out lending activities without limitations in the area of operations. Also, the banks will have to work towards achieving financial inclusion and 40 per cent of their lending should be towards the priority sector.	Can accept deposits of up to INR 0.1 million a customer and issue debit cards. It can also carry out payments and remittance services and is allowed to distribute insurance and mutual fund products. Payments banks can also serve as a business correspondent of another bank.	They will primarily undertake basic banking activities of accepting deposits and lending to un-served and under-served sections, including small business units, small and marginal farmers, micro and small industries and unorganized sector entities. There will not be any restriction in the area of operations of small finance banks.
Promoter's Contribution	The NOFHC and the bank will not have any exposure to the promoter group. The bank will not invest inequity/debt capital instruments of any financial entities held by the NOFHC.	The promoter's minimum initial contribution to the paid-up equity capital should be at least 40 per cent for the first five years from the start of its business.	The promoter's minimum initial contribution to the paid-up equity capital of small finance bank should be at least be 40 percent and needs to be gradually brought down to 26 per cent within 12 years from the start of operations.
Foreign Shareholding	The aggregate non-resident shareholding in the new bank will not exceed 49 per cent for the first five years, after which it will be according to the existing policy - 49 percent under the automatic route and 74 percent under the approval route.	The foreign shareholding in payments banks would be according to the foreign direct investment (FDI) policy for private sector banks - 49 per cent under the automatic route and 74 percent under the approval route.	Foreign shareholding in small finance banks would be according to the FDI policy for private sector banks- 49 per cent under the automatic route and 74 per cent under the approval route.
Other Conditions	The bank's board should have a majority of independent directors. It needs to open at least 25 per cent of its branches in unbanked rural centres (population of up to 9,999, according to the latest census). Also, banks promoted by groups having 40 per cent or more assets/income from non-financial business will require RBI's prior approval for raising paid-up voting equity capital beyond INR 10 billion or for every block of INR 5 billion.	The operations of the bank should be fully networked and technology-driven from the beginning and it should also have a high powered customer grievances cell to handle complaints.	The small finance bank will be subject to all prudential norms and regulations of RBI, as applicable to existing commercial banks, including requirement of maintenance of cash reserve ratio and statutory liquidity ratio. Apart from this, they will be required to extend 75 per cent of adjusted net bank credit to the priority sector. Also, at least 50 per cent of its loan portfolio should comprise loans and advances of up to INR 2.5 million.

One strand of the literature on Indian banking has focused on the impact of regulatory policy changes in efficiency. Bhattacharyya et al. (1997) studied public sector banks during the pre-reform period, i.e. 1970-1992 and found that the productivity growth for these banks were curtailed by nationalization. Kumbhakar and Sarkar (2003) expanded on this study by considering both public and private sector banks but for the period 1985-1996. They incorporate the distortionary effect of deregulation as measured changes in input prices. They do not find evidence for significant improvement in total factor productivity growth for both private and public sector banks after the reforms relative to the pre-reform period. They find private sector banks to have higher total factor productivity relative to public sector banks over the entire sample period. Casu et al. (2013) use a meta-frontier approach and find increasing productivity growth among Indian banks between 1992-2009. They find differences in productivity gains across ownership structures. They find foreign banks are found to have the most efficient production technology which is subsequently adopted by the remaining banks in the market. Gulati and Kumar (2016) also use a meta-frontier approach to study the impact of global financial crisis on profit efficiency of Indian banks. Their study covers the period from 2003-04 to 2012-03. They find a mild decline in profit efficiency of banks post the global financial crisis. Similar to Casu et al. (2013), they also find differential impact on efficiency by ownership type of banks. While new private banks experienced largest decline in profit efficiency, they find evidence that foreign banks continue to employ the most efficient production technology that is subsequently adopted by rest of the market. Combining the results from these two studies, one can see that the regulatory reforms initiated in the 1990?s has led to an overall increase in efficiency of Indian banks. However, Sengupta and Vardhan (2020) find that efficiency gains stagnated since 2011-12 in the Indian banking industry. Using the Malmquist Index, they study 33 scheduled commercial banks between 2002-2018 and find that increasing non-performing loans on banks' balance sheets has led to a decline in productivity growth in the banking sector.

The second strand of literature on the Indian banking industry seek to measure competi-

tion and changes in market structure over time. Prasad and Ghosh (2007) compute the Panzar-Rose H-statistic[6] for the period 1996-2004 and find that the Indian banking industry is characterized by monopolistic competition. They compute the H-Statistic using both interest revenue and total revenue as the dependent variable. They argue that this is evidence for the market being contestable where the incumbent firms are pricing close to competitive level because of threat of potential competition. However, they do not test whether the market is in long-run equilibrium which is necessary to validate the estimated results[7]. Perera et al. (2006) computes the Panzar-Rosse H-statistic for four countries in the Asian subcontinent - Bangladesh, India, Pakistan and Srilanka. For the time period, 1995-2003, they find monopolistic competition in the banking industry in Bangladesh, India, Pakistan and Srilanka. In India, they find that banks are efficiently moving away from labor intensive process to capital intensive ones as indicated by positive relationship between capital costs and bank revenue. They also find that in India and Srilanka, the traditional interest-based banking market is less competitive than the overall market. This could be because of the presence of foreign banks who have more fee and commission based products [8].

Another indirect way to measure competition in the banking industry can be found in Sensarma (2008). The author uses a stochastic frontier analysis to measure profit efficiency of Indian banks for the period 1986-2005. The author finds that profit efficiency and profit productivity declined over time despite the reforms in the 1990?s. This is interpreted as a result of increased competition in the industry as increased competition is leading to lower profits. Zhao et al. (2010) uses a stochastic cost frontier model to estimate marginal cost of individual banks and then use the estimated marginal costs to test for persistence of profits in the Indian banking industry. They study the period 1992-2004, which is divided into

[6]Panzar-Rose H-statistic measures the sum of elasticities of the price of output to changes in price of inputs.

[7]The long-run equilibrium test for the market was introduced by Bikker and Haaf (2002)

[8]On the other hand, in Bangladesh and Pakistan, they find that the traditional interest-based banking market is more competitive than the overall market (which includes fee and commission based banking products).

pre and post periods (1992-1997 and 1998-2004). They find that the regulatory reforms in 1992 and 1998 has led to decreasing markup's in the market for loans in the post period (1998-2004) relative to the pre-period (1992-1997). That is, they find evidence of increasing competition in the market. Their estimates can be interpreted in a similar way as decreasing Lerner index for the market for loans. The line of reasoning in this strand of literature is that increases in efficiency is due to increasing competition as competitive pressures incentivise banks to undertake measures to improve their cost efficiency and profit efficiency.

This study adds to the literature in two ways. First, I evaluate the typical measures of competition, Lerner Index and Boone indicator, and find that these measures are sensitive to measurement errors. These measures are also sensitive to changes in policies related to bad loans and its realization on bank's balance sheets. Second, I show that the degree of competition in the Indian banking industry has actually decreased over time. This is in sharp contrast to evidence from existing studies that claim that the Indian banking industry is competitive. My findings align with Das and Kumbhakar (2016). These findings raise important questions regarding the existing measures of competition as well as the overall banking policies pursued. I use the findings in this chapter to validate the findings from a structural model of competition discussed in Chapter 2.

In the next section, I describe the two standard measures of competition - Lerner Index and Boone Indicator. I then discuss the results of these measures and how they vary over time within the Indian Banking industry.

1.2 Measures of Competition

I study two measures of competition: (a) Lerner index, and (b) Boone Indicator. Computing both Lerner Index and Boone indicator first requires estimation of the marginal cost. As

my study focuses on the market structure for loans, the estimated marginal cost is with respect to an additional unit of loan disbursed. I estimate the marginal cost using a translog cost function using individual bank-level data between the financial years, 2003-2021. The translog cost function has the following form:

$$ln(VC_{it}) = \alpha_0 + \Sigma_{g=1,2,3}\beta_g d_g + \Sigma_{t=1,...,T-1}\tau_t d_t + \Sigma_{i=1}^{6}\delta_i ln(x_i) \tag{1.1}$$

$$+\Sigma_{i=1,2,3}\Sigma_{j=1,2,3}\gamma_{ij} ln(x_i) * ln(x_j) + \eta Z_i t + \nu_{it}$$

This is the translog cost function for bank i in year t. The above specification includes three outputs and three inputs. The outputs are based on the assumption that any bank has three channels to earn income. The first channel is through interest income that they earn on the credit or loans disbursed. The second channel is through the returns on their investments. This channel is relevant in the context of banks in India, as banks are required under regulations to hold certain proportion of their deposits in the form predetermined/pre-specified assets. The third and final channel is based on the commission and fees banks earn through activities other than disbursal of credit. The three inputs used seek to proxy for the cost of labor, physical capital, and funds required for the bank's operations. Specifically, each variable in the translog cost function is defined as:

- VC_{it} is the variable cost of production for bank i in period t. It is defined as operating expenses, which is the sum of interest expenses, expenses on employees, and other non-interest, non-employee related expenses.
- $X_{1\times 6} = [q_1, q_2, q_3, i_1, i_2, i_3]$ is a vector of 3 output and 3 input variables.
- q_1 is the total amount of outstanding loans disbursed as at the end of the financial year, scaled by total assets of the bank, i.e. $q_1 = \frac{\text{Loans and Advances}}{\text{Total Assets}}$.
- q_2 is the total amount of securities for bank i and proxies for returns from

- q_3 is .This captures the non-interest income that the bank earns during the financial year. T

- i_1 measures price of labor. This is defined as personnel expenses per employee, i.e. $\frac{\text{Personnel Expenses}}{\text{Total Employees}}$.

- i_2 measures price of other inputs, defined as $\frac{\text{Other non-personnel and non-interest expenses}}{\text{Fixed Assets}}$.

- i_3 measures the price of funds, defined as $\frac{\text{Interest expenses}}{\text{Total Deposits + Money Market Borrowings}}$.

- d_g are dummy variables for bank's group (or bank type), where we consider three groups of banks - Public sector banks, Private sector banks, and Foreign banks[9]. I add these dummy variables because each bank group face slightly different sets of regulations and incentives in terms of their operations[10].

- d_t are dummies for year which are included to capture any unobserved factors that affects all banks.

- Z_{it} is the equity-to-assets ratio to factor for differences in risks in loan portfolios across banks (Berger and Mester (1997), Van Leuvensteijn et al. (2011)).

- ν_{it} is the error term.

This specification of the translog cost function is based on Van Leuvensteijn et al. (2011), where a bank's technology is assumed to be defined by a multiproduct production function. Under assumptions of linear homogeneity in inputs prices and cost-exhaustion, we can derive a dual cost function from this production function where output levels and input prices are used as arguments. The specification above is a second-order Taylor expansion around the mean of a generic dual cost function. All the variables - variable cost, outputs, and input prices - appear as natural logarithms. We apply assumptions of linear homogeneity in input

[9]See appendix A for more details about each bank group.

[10]Zhao et al. (2010) . Also, Casu et al. (2013) and Gulati and Kumar (2016) show that foreign banks were more efficient .

prices and cost exhaustion on the cost function by imposing the following restrictions on the parameters of the cost function (i=1,2,3 refers to outputs & i=4,5,6 refers to the three inputs). Assumption of cost exhaustion is :

$$\delta_4 + \delta_5 + \delta_6 = 1$$

This also reflects that the sum of costs of individual inputs must add up to the total variable cost. The assumption of linear homogeneity in input prices is imposed as:

$$\gamma_{4k} + \gamma_{5k} + \gamma_{6k} = 0; \text{where k=4,5,6}$$

$$\gamma_{j4} + \gamma_{j5} + \gamma_{j6} = 0; \text{where j=1,2,3.}$$

Linear homogeneity in input prices also imply cost exhaustion while at the same time it the squared and cross-product terms of inputs and outputs must add up to zero. In addition to these two conditions, the condition of symmetry is also imposed, i.e. $\gamma_{jk} = \gamma_{kj}$ for all $j, k = 1, ..., 6$. With the imposition of the restrictions on the parameters, we estimate the following form of the translog cost function:

$$\widetilde{ln(VC_{it})} = \alpha_0 + \Sigma_{g=1,2,3}\beta_g d_g + \Sigma_{t=1,...,T-1}\delta_t d_t$$

$$+\delta_1 ln(q_1) + \delta_2 ln(q_2) + \delta_3 ln(q_3) + \delta_4 ln(\widetilde{i_4}) + \delta_5 ln(\widetilde{i_5})$$

$$+\gamma_{11}(ln(q_1))^2 + \gamma_{22}(ln(q_2))^2 + \gamma_{33}(ln(q_3))^2 + \gamma_{44}(ln(\widetilde{i_4}))^2 + \gamma_{55}(ln(q_5))^2$$

$$+\gamma_{12}(ln(q_1) * ln(q_2)) + \gamma_{13}(ln(q_1) * ln(q_3)) + \gamma_{23}(ln(q_2) * ln(q_3)) + \gamma_{45}(ln(\widetilde{i_4}) * ln(\widetilde{i_5}))+$$

$$\gamma_{14}(ln(q_1) * ln(\widetilde{i_4})) + \gamma_{15}(ln(q_1) * ln(\widetilde{i_5})) + \gamma_{24}(ln(q_2) * ln(\widetilde{i_4}))$$

$$+\gamma_{25}(ln(q_2) * ln(\widetilde{i_5})) + \gamma_{34}(ln(q_3) * ln(\widetilde{i_4})) + \gamma_{35}(ln(q_3) * ln(\widetilde{i_5}))$$

where, $\widetilde{VC} = \frac{VC}{i_3}$, $\widetilde{i_1} = \frac{i_1}{i_3}$, $\widetilde{i_2} = \frac{i_2}{i_3}$. I divide the variable cost and the two input prices by the remaining input price to impose the assumptions of linear homogeneity and cost exhaustion. After estimating the translog function, we estimate the marginal cost of additional unit of loan disbursed as:

$$\widehat{MC} = \frac{\delta\widetilde{VC}}{\delta q_1} = (\frac{\widetilde{VC}}{q_1}) * \frac{\delta(ln(\widetilde{VC}))}{\delta ln(q_1)}$$

$$= \widehat{\delta_1} + [2 * \widehat{\gamma_{11}} * ln(q_1)] + [0.5 * \widehat{\gamma_{12}} * ln(q_2)] + [0.5 * \widehat{\gamma_{13}} * ln(q_3)] + [0.5 * \widehat{\gamma_{14}} * ln(\widetilde{i_4})] + [0.5 * \widehat{\gamma_{15}} * ln(\widetilde{i_5})]$$

To compute the empirical Lerner index, we also require the price of loans. I use two prices for this purpose: (a)the interest rate data on loans for each bank, and (b) implied price of loans[11]. The empirical Lerner Index is computed as:

$$\text{Lerner Index} = \frac{p - \widehat{mc}}{p}$$

Once the marginal cost is estimated, we then estimate the Boone indicator using the following model:

$$ln(ms_{it}) = \alpha + \beta ln(mc_{it}) \tag{1.2}$$

where, ms_{it} is the market share of bank i of total outstanding loans in the banking industry at the end of year t[12]. Here, the parameter β is the Boone indicator and apriori, β is expected to be negative, i.e. firms with higher marginal costs are expected to have smaller market shares. More efficient banks that have smaller marginal costs are expected to gain market share at the expense of less efficient banks (that have higher marginal costs). As the degree of competition increases, the change in market share is expected to be stronger. Therefore,

[11]This is computed as the interest earned on all loans and advances divided by the total amount of outstanding loans and advances.

[12]The empirical model is specified in log-linear terms to address heteroscedasticity. This specification also allows one to interpret β as a measure of elasticity and compare across specifications.

greater the competition, greater the value of $|\beta|$, where $\beta < 0$. One of the limitations of the Boone indicator model is that we can very well observe $\beta > 0$. This is because differences in the product quality and design is ignored within this setup. In the context of Indian banks, a potential issue in ignoring product quality and design is related to bank ownership. As the government is the major stakeholder in public sector banks, there is an implicit guarantee that the government will not allow such banks to fail [13]. The Boone indicator model also assumes that banks, in general, pass on some of the efficiency gains to customers (borrowers) by lowering the price of loans. There is another implicit assumption that banks have to provide similar levels of quality of service over time and that they assimilate each other's best practices, especially in terms of innovations leading to efficiency gains. Boone indicator, however, allows us to focus on competition within submarkets rather than the banking industry as a whole.

1.3 Data Sources and Summary Statistics

The data used for this analysis is compiled from multiple sources. The two primary sources of data are from the Database on Indian Economy (DBIE), published by the Reserve Bank of India and annual "Performance Highlights" reports published by the Indian Banks' Association (IBA). The latter was used to complement and fill gaps in data from the former source when necessary. The data is published according to the Indian financial calendar where the financial year spans from the month April of a given year to March of the next year.

Data can be classified into the following categories based on their source level of disaggregation:

[13] This is quite evident in the series of amalgamations between public sector banks in the past decade where poorly performing banks were amalgamated with their larger counterparts under the direction of the government.

- Data from DBIE (RBI):
  - Annual financial statements of scheduled commercial banks for the period 2003-2021. This data is at the country-bank-year level.
  - Basic Statistical Returns of Scheduled Commercial Banks to compile data on credit and deposits. This data is at state-bank group-year level.
  - Bank Branch Statistics covering data on number of branches and number of new offices opened. The data is at the bank-district/state-quarter level covering 2005-2021.
- Data from IBA:
  - Annual performance highlights, which includes financial statements between 2003-2020. The data is at the country-bank-year level.
- Data from Ministry of Statistics and Program Implementation (MOSPI):
  - Index of Industrial Production - aggregate index and its components at a monthly frequency.
  - Consumer Price Index and Wholesale Price Index - aggregate index data at a monthly frequency.
- Data on interest rates charged by Public, Private and Foreign banks were collected from RBI's website through web-scraping.

Two other datasets used for this study is data on GDP Deflator which was downloaded from FRED database. I collected data on the daily yield on the 10-year benchmark bond from Bloomberg.

We have data on unique 141 scheduled commercial banks during the sample period spanning

from March, 2004 to March, 2021 [14] . Though certain data is available at the district-level and on a quarterly frequency, I have restricted the analysis to annual frequency and defined markets at the level of a state[15]. This is because all the variables used in the analysis are available at an annual frequency and at the state level. That is, data is not available for all variables at the district-level and on a quarterly basis to perform analysis at the disaggregated level and frequency. Additionally, banks have merged with each other or certain public sector banks have been amalgamated into larger public sector banks during the sample period. I have considered each bank as separate entity until the date of their merger/amalgamation. There are also two instances were a state was split into multiple parts during the sample period [16] As data is not available at disaggregated level for these newly created states or administrative regions, I have defined them to be a single market for the duration of the analysis. Appendix A provides more details of the banking industry in India.

For the bank-level analysis, we consider all banks - banks that operate in multiple markets as well as single market banks - the analysis[17]. The final dataset consists of 110 unique banks spanning the years 2005-2021 over 34 different markets (states). The panel is not balanced as some banks exit the industry or merge or be amalgamated with other banks during the sample period. All amounts have been deflated using GDP Deflator and are defined in terms of 2011-12 Indian Rupee (INR).

Tables 1 and 2 list the summary statistics of the key variables used in the bank-level model used to test the mutual forbearance hypothesis in the presence of multimarket contact.

[14]Even though financial data is available for years prior to 2005, data on other variables are not available and cannot be imputed from available data. For instance, data on branches by state (or district) for each bank is available only starting March, 2005. However, I am able to compute the same data for 2004 using related data on branch openings that covers the financial year 2004.

[15]States represent the higher level of administration above districts

[16]A new state of Telegana was created out of the existing state of Andhra Pradesh in 2014. Similarly, the state of Jammu and Kashmir was split into two union territories - Jammu and Kashmir, and Ladhakh.

[17]As robustness check, we consider the subset of public and private banks only for the analysis. The results are reported in Appendix B and they indicate evidence similar to the results with all banks included. I could not run separate models for each bank group because of lack of degrees of freedom given the number of fixed effects and parameters to be estimated.

Table 1.2: Bank-wise summary statistics of key variables

Variable Name	N	Mean	Standard Deviation	Median	Min	Max
Total Earnings (INR Million)	1386	91935.12	196821.88	25690.93	61.44	2295241.25
Total Assets (INR Billion)	1386	1102.91	2467.63	306.69	0.43	30964.99
Return on Assets (%)	1386	8.60	2.80	8.35	0.47	35.81
Bad Loans (INR Million)	1386	17569.92	51669.08	1279.63	0.00	873606.31
Total Loans (INR Million)	1386	642330.50	1470431.88	155704.22	1.82	17640678.00
Equity (INR Billion)	1386	83.23	174.91	29.73	0.07	1760.11
Total Number of Branches	1386	1159.87	2342.07	246.00	1.00	24949.00
Operating Expenses (INR Million)	1386	19509.42	44895.88	5825.86	29.81	570301.00

Table 2.1 reports the summary statistics used for constructing the Lerner index and Boone indicator. These summary statistics have been averaged across all banks and across all years. The mean and the standard deviation of these variables indicate a lot of variation across banks and across time. All rupee variables have been deflated using the GDP Deflator and are measured in 2011-12 Indian Rupee (INR). I have also scaled the variables into billions or millions depending on the distribution of its values. Total number of branches is represented in actual units. All the statistics reported are for public sector banks, private sector banks, and foreign banks.

Table 2.2 describes the summary statistics of all state level variables used in the bank-level measures of competition. Similar to the variables in Table 2.1, all rupee denominated variables have been deflated and converted to 2011-12 INR using the GDP deflator. To apply the bank-level analysis, each of these variables have been converted to level of the bank by taking weighted sum of each of these state-level variable for every bank in the sample. The weights are defined as the market share of the given bank in the given year in each state, where the market share are computed using the branches [18]. I have deflated each of the rupee denominated variable by the GDP deflator to define all these variables in terms of 2011-12 INR (same as in Table 2.1). The average interest rate on loans during the sample

[18]I do not have access to any other bank-wise state-level variable such as loans, deposits, assets etc that can be used instead.

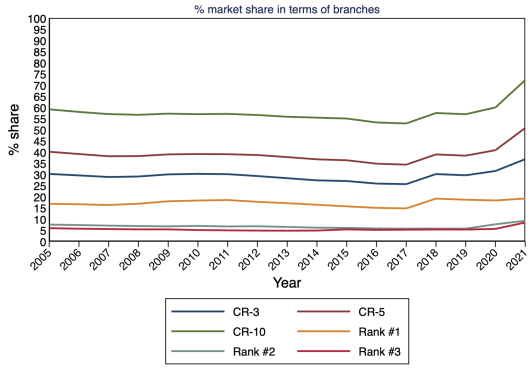
Table 1.3: Summary statistics for state-level features

Variable	N	Mean	Standard Deviation	Median	Min	Max
State HHI	577	0.16	0.14	0.10	0.04	0.82
State Population (millions)	577	38.50	45.60	25.53	0.00	231.14
Total Branches in State	577	2786.82	3178.32	1503.00	10.00	13794.00
Total Banks in State	577	38.06	16.38	38.00	2.00	82.00
Total Number of Credit Accounts (millions)	577	3.78	6.65	1.14	0.00	54.94
Growth rate in Credit Accounts (%)	577	10.92	15.00	8.14	-35.71	143.33
Total Amount of All Credit (INR Billion)	577	1468.38	2989.39	331.35	0.23	21604.10
Growth rate in All Credit (%)	577	11.60	11.66	10.07	-20.81	104.62
Total Number of Small Borrower Credit Accounts (millions)	577	2.95	5.35	0.85	0.00	42.54
Growth in Number of Small Borrower Credit Accounts (%)	577	9.55	20.47	5.86	-50.00	178.62
Total Credit to Small Borrowers (INR Billion)	577	121.49	187.90	39.76	0.13	1066.22
Growth in Credit to Small Borrowers (%)	577	4.99	16.26	2.96	-59.06	100.01
Total Number of Deposit Accounts (millions)	577	30.79	41.97	13.11	0.03	217.36
Growth in Number of Deposit Accounts (%)	577	9.90	7.55	9.77	-18.16	52.51
Amount of Total Deposits (INR Billion)	577	1930.93	3168.51	697.74	2.02	21379.99
Growth in Total Deposits (%)	577	9.33	7.84	8.82	-19.48	50.66
Total Number of Current Deposit Accounts (millions)	577	1.36	1.93	0.60	0.00	14.47
Growth in Number of Current Deposit Accounts (%)	577	13.30	26.12	9.38	-74.85	199.91
Total Amount of Current Deposits (INR Billion)	577	203.98	407.72	69.45	0.08	3233.40
Growth in Current Deposits (%)	577	10.10	28.89	6.74	-72.23	264.79
Total Number of Savings Deposit Accounts (millions)	577	24.35	34.51	9.98	0.03	184.74
Growth in Number of Savings Deposit Accounts (%)	577	11.24	8.59	10.85	-25.22	77.57
Total Amount of Savings Deposits (INR Billion)	577	549.68	730.46	262.97	1.19	4912.68
Growth in Savings Deposits (%)	577	549.68	730.46	262.97	1.19	4912.68
Total Number of Term Deposit Accounts (millions)	577	5.08	6.15	2.62	0.00	30.75
Growth in Number of Term Deposit Accounts (%)	577	4.99	10.35	5.31	-51.62	89.52
Total Amount of Term Deposits (INR Billion)	577	1177.27	2142.18	372.51	0.63	14071.91
Growth in Term Deposits (%)	577	8.94	13.69	8.18	-44.99	177.10
Per capita number of deposit accounts in state (millions)	577	0.98	0.74	0.79	0.10	3.87
Per capita number of credit accounts in state (millions)	577	0.10	0.08	0.07	0.02	0.48
Per capita number of small borrower credit accounts in state (millions)	577	0.08	0.06	0.05	0.01	0.39
Per capita number of current deposit accounts in state (millions)	577	0.05	0.05	0.03	0.00	0.27
Per capita number of savings deposit accounts in state (millions)	577	0.74	0.51	0.63	0.08	2.58
Per capita number of term deposit accounts in state (millions)	577	0.19	0.22	0.13	0.01	1.30
Per capita amount of deposits in state (INR Thousands)	577	76.18	99.65	43.36	6.18	478.37
Per capita amount of credit in state (INR Thousands)	577	51.76	96.80	20.42	1.66	516.18
Per capita amount of credit to small borrowers in state (INR Thousands)	577	3.52	2.68	2.52	0.84	15.01
Per capita amount of current deposit in state (INR Thousands)	577	7.60	10.43	4.31	0.69	79.25
Per capita amount of savings deposit in state (INR Thousands)	577	21.78	22.08	14.98	2.29	145.83
Per capita amount of term deposit in state (INR Thousands)	577	47.03	69.85	22.42	1.94	341.69

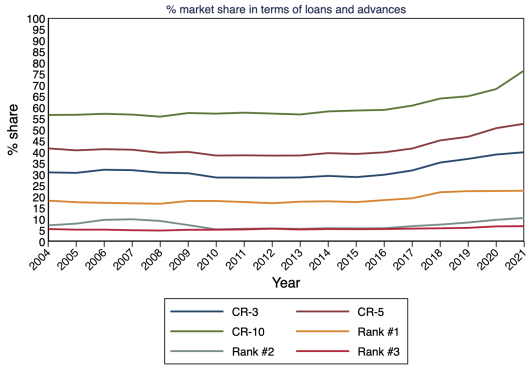
period is 10% per annum (after rounding)[19].

I also plot the 3-firm, 5-firm and 10-firm concentration ratios each year in terms of number of active branches, loans and advances, deposits, and total assets of public, private and foreign banks in Figure 2.2 . All four sub-figures show that there is a dominant firm in the market which has a market share nearly double that of the next largest bank in the industry. The dominant bank discussed here is the State Bank of India, a public sector bank, and it maintains its dominance in market share throughout the entire sample period. Moreover, we observe an increase in concentration ratio towards the end of our sample period.

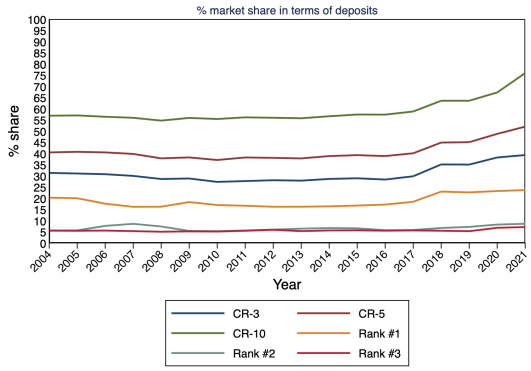
[19]This is not the imputed interest rate on loans as it can be biased by the presence of non-performing loans which became a severe issue during the latter half of the sample period.



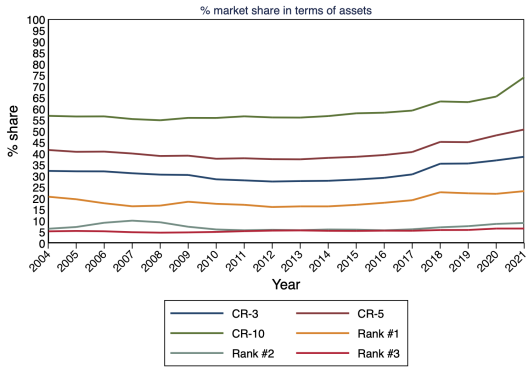
(a) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of number of branches, 2005-2021



(b) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of loans and advances, 2004-2021



(c) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of deposits, 2004-2021



(d) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of total assets, 2004-2021

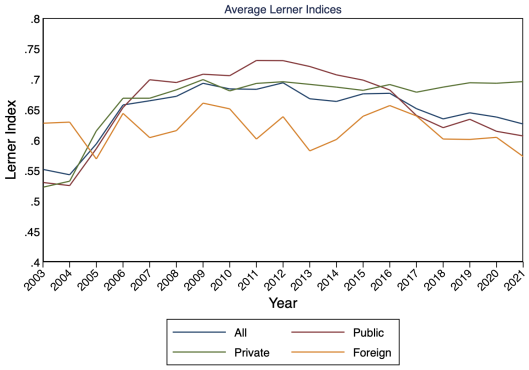
Figure 1.1: Market shares and concentration ratios - 2004-2021

From Figures 2.2, we can see that concentration ratios have increased since 2012. This is the case irrespective of whether we compute market shares in terms of branches or outstanding loans or deposits or total assets of banks. Moreover, we can also see that the top bank's share is increasing towards the end of the sample. The bank with the highest market share throughout the whole sample period is State Bank of India (SBI), a public sector bank. It's market share is increasing towards the end of the sample period as a few other smaller public sector banks have been amalgamated with it.

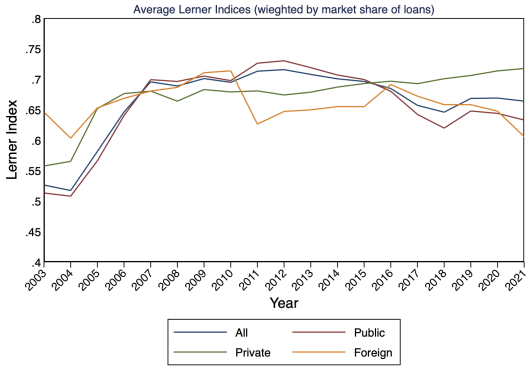
1.4 Results

I start by discussing the estimates for the Lerner Index and the Boone indicator for the period under review, 2004-2021.

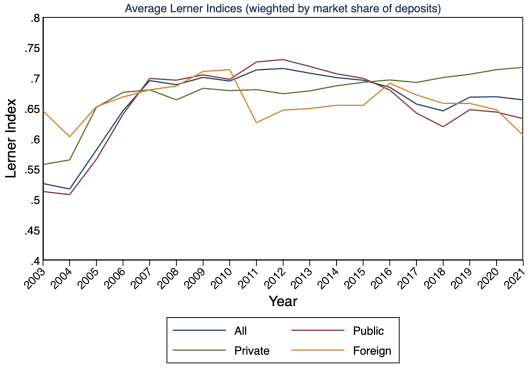
Figure 1.2: Estimated Lerner Index (based on imputed price of loans)



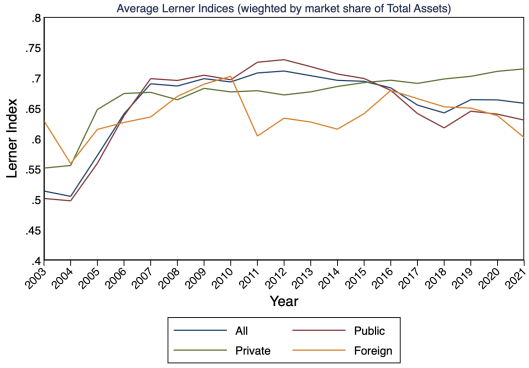
(a) Unweighted



(b) Weighted - by loans of banks



(c) Weighted - by deposits of banks



(d) Weighted - by total assets of banks

Based on the estimation of the translog cost function in 1.1, I compute the Lerner index using the imputed price of loans from the annual balance sheets of banks. The Lerner index represented in Figure 1.2 measures the degree of competition in the market for loans. I then aggregate it year-wise to determine how the Lerner index has changed over time. Higher values of the index indicate decreasing competition while lower values indicate increasing

competition as the price is pushed closer to the competitive equilibrium price. Once I have estimated 1.1, I use the estimates to compute the Lerner index for each individual bank. This then allows me to plot the Lerner index averaged over all banks or over all banks within a specific bank group. Figure 1.2 plots these average yearly Lerner indices. To ensure representativeness while aggregating the estimated Lerner indices, I have weighted them using the bank's share of loans and advances, deposits, and their total assets. Subplots (a), (b), (c), and (d) and plot the aggregated Lerner indices for each year[20].

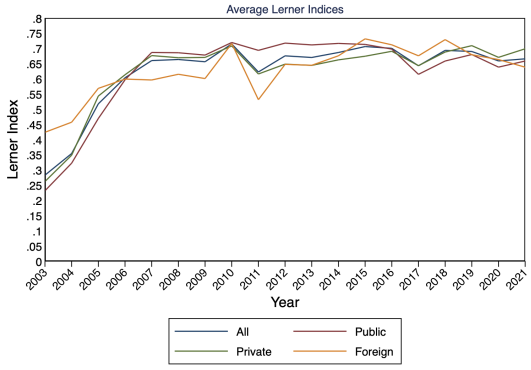
Figure 1.2 shows evidence of decreasing competitive pressures in the market for loans at the beginning of the study period. This is true across all subplots. We find that the Lerner indices remain at their higher levels for most of the sample period for the overall banking industry. For foreign banks, there is a reduction in the Lerner indices between 2009 and 2012. This is quite likely a result of the Global Financial crisis. We observe that the Lerner indices for foreign banks stabilize around the lower level for the rest of the sample period. The Lerner index for the whole industry and for public sector banks decreases around 2016. However, the Lerner index for private sector banks remains high and even increases between 2016 and 2021. This is puzzling as I would expect the opposite pattern in Lerner indices. I expected the Lerner index for private banks decreases due to the increase in competition from entry of two new private sector banks in 2014. On the other hand, the Lerner index for public sector banks should decrease given that a number of smaller public sector banks were amalgamated with their larger counterparts, thus decreasing competition within this group of banks. A potential explanation for the decrease in Lerner index is related to its construction. I have used the imputed price loans as the price of loans to compute the index. The imputed price of loans is the ratio of interest earned on loans and advances over the total amount of loans and advances. However, during this particular period, public sector banks suffered from increasing amounts of non-performing loans. This can then lead to a

[20]I have used within bank group weights of each bank (defined by their shares of loans, deposits and total assets among banks within the same bank group) while aggregating the Lerner index by bank group.

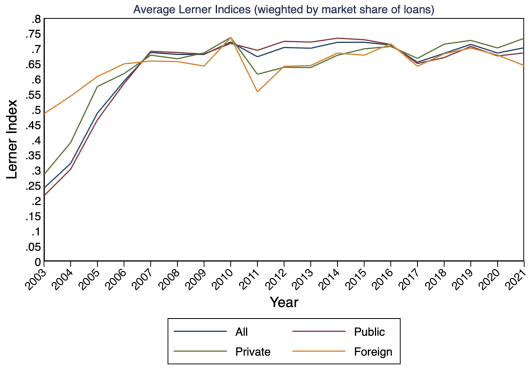
decrease in interest earned on loans and advances and thus create a lower price of loans even though these banks might still be charging a higher interest rate for their loans.

To address this shortcoming, I have re-calculated the Lerner index using a bank's base rates (or prime lending rates) as the price of loans [21]. In Figure 1.4, I plot the Lerner indices as before (unweighted version and weighted versions). We can now see that there is no significant decline in the Lerner index after the initial increases at the beginning of the sample period and remains at the higher level for the rest of the sample period. The Lerner indicator implies decreasing competition in the banking industry during the sample period.

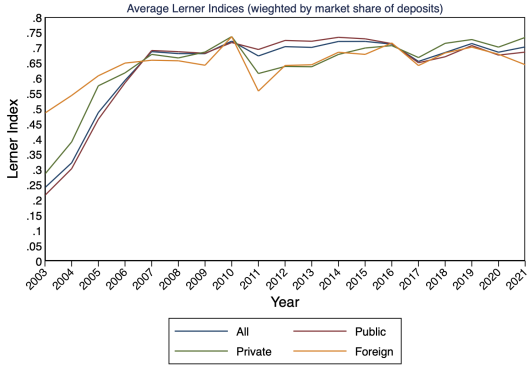
Figure 1.3: Estimated Lerner Index (based on bank's base rate for loans)



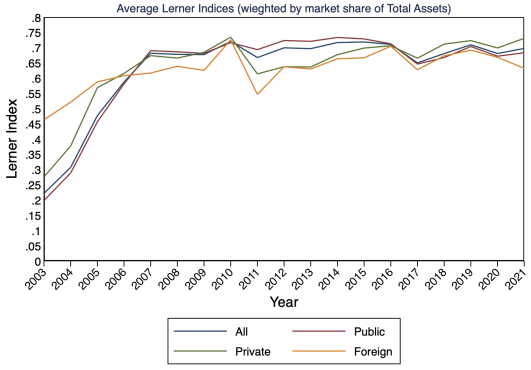
(a) Unweighted



(b) Weighted - by loans of banks



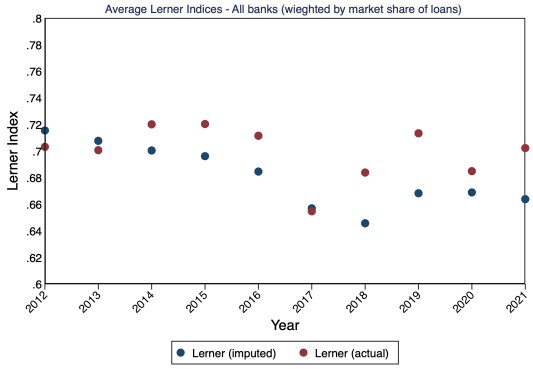
(c) Weighted - by deposits of banks



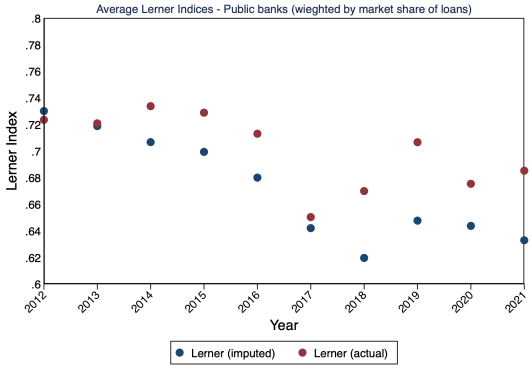
(d) Weighted - by total assets of banks

[21] These interest rates on loans are the base rates that a bank would charge a client with zero credit risk. As I do not have access to individual loan level data, this is the best proxy available for price of loans at the bank level.

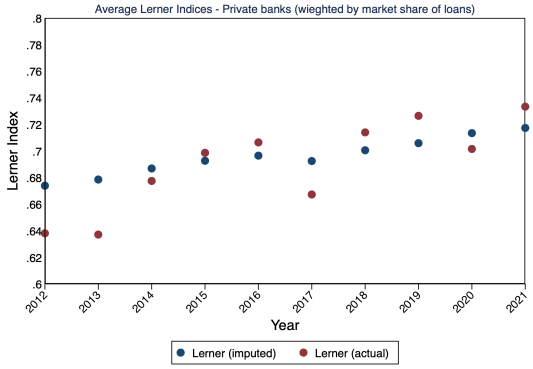
Figure 1.4: Comparing estimated Lerner Index (based on imputed price of loans and bank's base rate for loans)



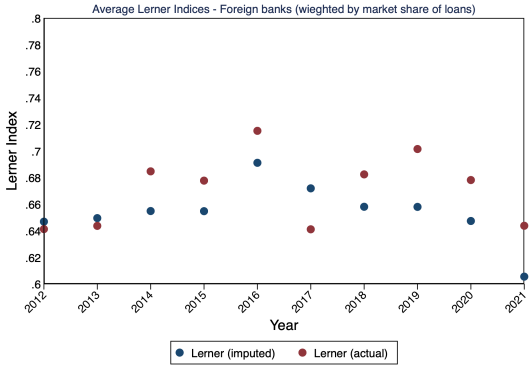
(a) Avg estimated Lerner index (weighted by market share of loans) for all banks - 2012-2021



(b) Avg estimated Lerner index (weighted by market share of loans) for Public banks - 2012-2021

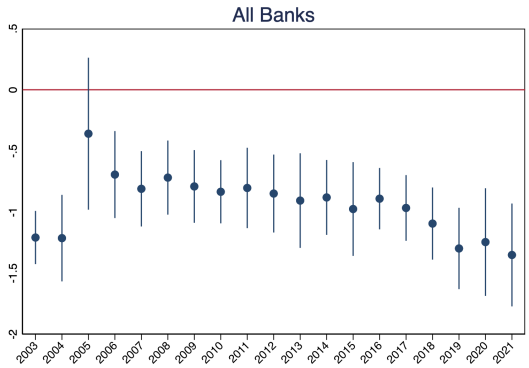


(c) Avg estimated Lerner index (weighted by market share of loans) for Private banks - 2012-2021

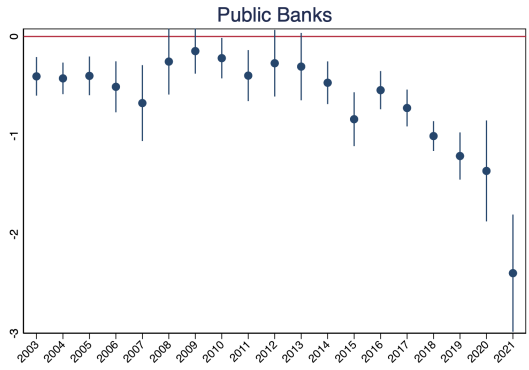


(d) Avg estimated Lerner index (weighted by market share of loans) for Foreign banks - 2012-2021

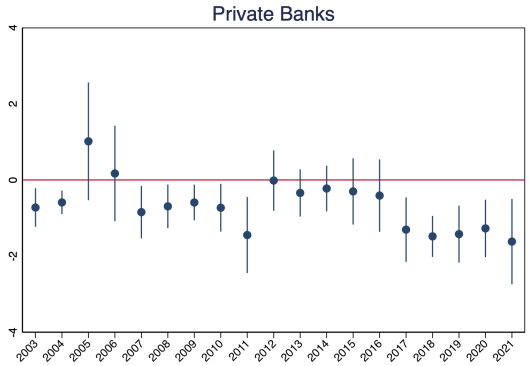
Figure 1.5: Year wise estimated Boone Indicator



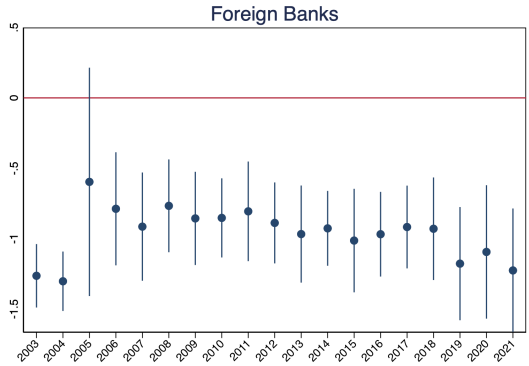
(a) Year wise Boone indicator - All banks



(b) Year wise Boone indicator - Public Sector banks



(c) Year wise Boone indicator - Private Sector banks



(d) Year wise Boone indicator - Foreign banks

Figure 1.5 plots the Boone indicator along with their 95% confidence interval. These estimates are also reported in Table 1.4. Panel A of Figure 1.5 (Column 1 of Table 1.4) plots the Boone indicator,β , from eq 1.2 , after estimating the marginal cost via the translog cost function. Panel B of Figure 1.5 (Column 2 of Table 1.4) reports the Boone indicator for the set of Public Sector Banks only while Panel C of Figure 1.5 (Column 3 of Table 1.4) reports the Boone indicator for Private Sector Banks, and Panel B of Figure (Column 2 of Table 1.4) plots the Boone indicator for Foreign banks. The difference in measure of competition using the Boone indicator relative to the Lerner index could be due to the definition of the dependent variable used to estimate the Boone indicator. I use the market share of loans as

the dependent variable while the Lerner index is indirectly measuring the return on loans made by the bank.

Table 1.4: Boone Indicator (yearwise)

	(1)	(2)	(3)	(4)
	All	Public	Private	Foreign
2003	-1.209***	-0.403***	-0.723**	-1.264***
	(0.131)	(0.115)	(0.298)	(0.135)
2004	-1.214***	-0.423***	-0.589***	-1.303***
	(0.213)	(0.0937)	(0.181)	(0.126)
2005	-0.359	-0.398***	1.014	-0.597
	(0.375)	(0.115)	(0.911)	(0.484)
2006	-0.694***	-0.509***	0.170	-0.788***
	(0.214)	(0.151)	(0.737)	(0.240)
2007	-0.810***	-0.673***	-0.844**	-0.915***
	(0.186)	(0.225)	(0.406)	(0.230)
2008	-0.719***	-0.254	-0.692**	-0.767***
	(0.183)	(0.196)	(0.336)	(0.197)
2009	-0.791***	-0.148	-0.591**	-0.856***
	(0.179)	(0.134)	(0.275)	(0.198)
2010	-0.834***	-0.219*	-0.729*	-0.853***
	(0.156)	(0.120)	(0.369)	(0.168)
2011	-0.803***	-0.395**	-1.445**	-0.806***
	(0.198)	(0.152)	(0.588)	(0.211)
2012	-0.849***	-0.270	-0.0154	-0.888***
	(0.192)	(0.198)	(0.467)	(0.172)
2013	-0.907***	-0.304	-0.340	-0.968***
	(0.233)	(0.200)	(0.365)	(0.206)
2014	-0.881***	-0.467***	-0.224	-0.927***
	(0.185)	(0.127)	(0.352)	(0.159)
2015	-0.975***	-0.836***	-0.301	-1.014***
	(0.231)	(0.160)	(0.512)	(0.220)
2016	-0.891***	-0.542***	-0.410	-0.969***
	(0.151)	(0.113)	(0.560)	(0.180)
2017	-0.967***	-0.723***	-1.303**	-0.917***
	(0.162)	(0.109)	(0.497)	(0.176)
2018	-1.095***	-1.006***	-1.480***	-0.930***
	(0.178)	(0.0887)	(0.317)	(0.217)
2019	-1.298***	-1.208***	-1.421***	-1.178***
	(0.201)	(0.140)	(0.441)	(0.240)
2020	-1.246***	-1.358***	-1.271***	-1.094***
	(0.265)	(0.300)	(0.443)	(0.283)
2021	-1.351***	-2.391***	-1.617**	-1.226***
	(0.254)	(0.348)	(0.661)	(0.263)
Observations	1333	456	392	485
Within R-Squared	0.514	0.703	0.384	0.640
Between R-Squared	0.0327	0.585	0.301	0.0522
R-Squared	0.514	0.703	0.384	0.640
Overall R-Squared	0.0572	0.192	0.163	0.118

The estimated coefficients for the Boone indicator for the overall market show evidence of increasing competition over time as the absolute magnitude of the Boone indicator is increasing over time. This is contrary to what we observe with the Lerner index. Breaking down the Boone Indicator by bank group, we see that the changes in competitive pressures differ across bank groups. For foreign bank, after the initial decrease in the Boone indicator (in absolute terms) between 2003 and 2005, the Boone indicator then decreases over the rest of the sample period. For private banks on the other hand, the Boone indicator is not always negative (as predicted by theory). However, it is statistically less than zero for most years in the sample and the indicator shows an increase in competition for private sector banks towards the end of the sample period (2016-2021). A similar comment can also be made for Public sector banks, as the Boone indicator initially decreases (in absolute terms) between 2007 and 2008 and then increases over the rest of the sample period. There is a sharp drop in the Boone indicator for Public sector banks between 2016-2021.

1.5 Conclusion

Why do the Boone indicator and Lerner indices show different results? First, we have to keep in mind that these indicates are based on partial equilibrium analysis and requires computation of marginal costs. Neither are measures constructed based on general equilibrium conditions of the market. We need an measure that simultaneously estimates both demand as well as the profit maximizing equilibrium conditions. In case of the Lerner index, the price is an imputed price which ignores factors that can simultaneously affect demand for loans as well the firm's ability to price their output. One such factor is the presence of non-performing loans that can increase the risk faced by banks in the output market. This not only raises the bank's costs, it also simultaneously leads to a lower imputed price as interest income on loans fall as non-performing loans increase. I show in this study that addressing this limitation leads to an alternative conclusion by using a direct measure of the

interest rates charged by banks for their loans and advances. Boone indicator suffers from a similar issue as it is also measuring marginal costs and its relationship with market shares of individual banks in a partial equilibrium framework.

Second, any cost function estimates can suffer from bias and consistency if there is input misallocation (Kumbhakar and Wang (2006). These measures are most likely to be plagued by allocative errors in inputs and therefore, estimates of the marginal cost used in construction of these measures. This is particularly relevant in the context of the Indian banking industry as the public sector banks are more likely to disburse credit or open new branches at the behest of the government rather than make these decisions with an objective to maximize profits or minimize costs (or both). The results from the Lerner indices that indicate decreasing competition in the banking market are consistent with the findings of Das and Kumbhakar (2016). Using an input distance function approach, they avoid the inconsistency in cost estimates even when input allocations are not optimal. Das and Kumbhakar (2016) find that larger banks, measured in terms of their total assets, enjoyed greater market power either due to cost advantages or due to their capacity to impose higher prices.

Burgess and Pande (2005a) and Young (2021) find positive effects of policies geared towards increasing branches in under-banked areas. However, continued mergers and amalgamations of banks can be counter-productive to branch expansion policies pursued by Reserve Bank of India and Government of India. If competition and contestability between banks is decreasing, the price of loans and advances will price out credit-seekers - both individuals and entrepreneurs - in under-banked areas. Decreasing competition can have a negative effect on access to formal finance and its use, especially for the minority communities. This includes women, households headed by women, individuals belonging to lower social groups (caste). While reverting to a fully controlled interest regime is not desirable, the findings in the study calls for greater attention of the policy makers to monitor how individual banks exert their market power in the market for credit.

Chapter 2

Multimarket contact and competition in the Indian Banking Industry

2.1 Introduction

The Reserve Bank of India, the monetary authority in India, introduced a policy change in 2005 whereby it incentivised banks to expand their branch network, especially in under-banked regions. The objective behind the policy was to improve access to banking services and finance in the country. In this study, I find evidence that the policy had the unintended effect of exacerbating the market power of banks, in turn allowing them to charge higher than competitive interest rates in the market for credit. I show that mulitmarket contact between banks in India has facilitated tacit collusion, consistent with the theory of mutual forbearance.

The banking industry in India offers a rich environment to study the impact of various policy changes on competition, efficiency and productivity. There has been quite a few policy changes since 1991 to increase competition in the Indian banking industry. These changes were introduced with the objective of improving the efficiency of banks in the industry, increasing access to finance and improve financial inclusion. In 1992, the monetary authority in India, the Reserve Bank of India (RBI) introduced changes in policy deregulating the market, particularly in terms of setting interest rates and allowing entry of private banks. In 1998, RBI introduced further reforms with the objective of strengthening financial stability. Another major change in policy occurred in 2005, where RBI made changes to how new

branch licenses were issued. This change was made to increase competition between banks especially in the under-banked regions of the country.

According to Structure-Conduct-Performance school of thought, collusion is easier in concentrated markets [1]. Therefore, easing entry and requirements for opening new branches is expected to reduce the market concentration and in turn, improve competition between banks. Even if market concentration does not decrease, easing entry and exit restrictions is expected to increase contestability in markets. However, expansion of bank's branch network also provides banks with a new channel to collude among themselves. Under the multimarket contact hypothesis, firms that operate in the same markets, either in terms of products or geography or both, have less incentive to compete aggressively against each other if they fear retaliation from rivals in other markets.

I find evidence the policy did result in facilitating collusion between banks. I measure the degree of multimarket contact between banks between 2005-2021 and find that their profits are positively and significantly associated with the levels of the bank's multimarket contact. Furthermore, I show, contrary to existing results, the equilibrium pricing behavior in the Indian banking market represents that of a cartel or a monopoly. I argue that the market is represents an oligopolistic market with a dominant firm that sets the market prices and rest of firms set prices off of this . I use a structural model developed by Bresnahan (1982) and Lau (1982) to identify the structure of the market. I use the structural model to show that, consistent with mutual forbearance theory, multimarket contact between banks has facilitated implicit collusion between banks in the market for loans.

This is the first study that examines multimarket contact in the Indian banking industry. This is also the first study that applies Bresnahan-Lau's general equilibrium approach to identify market structure parameters. My findings are in line with Das and Kumbhakar

[1] Bernheim and Whinston (1990), Feinberg (1984), Spagnolo (1999), Thomas and Willig (2006) provide theoretical framework supporting this hypothesis.

(2016) who find that the markups charged by Indian banks in the market for loans has increased since 2002. The evidence in this paper indicates that the access to finance is likely to have worsened as a result of the new regulations. Evidence suggest that larger banks have used the threat of retaliation in other markets to keep interest rates higher in underbanked market, thus defeating the original purpose of the policy.

2.2 Background and Summary of the literature

As a result of various regulatory policies pursued over the past 75 since its independence, the banking industry in India is characterized by a strong presence of public sector banks where the Government of India is the majority stakeholder. One of the major events was the nationalization of 14 scheduled commercial banks by the government of Indian 1969 and nationalization of another 6 private sector banks in 1980. The monetary policy authority, Reserve Bank of India (RBI), also imposed a wide range of price and quantity regulations that led to a scenario where large proportion of deposits were held in the form of reserves. A strictly administered interest rate regime was put in place by RBI and this created a high cost and low quality of financial intermediation services. In 1992, reforms were introduced following the recommendations of the Narasimhan Committee report of 1991. The government allowed for greater entry of privately owned banks as well as foreign banks to increase competition in the banking industry. In 1998, the interest rate regime was liberalized and government's stake in public sector banks were reduced substantially. Reforms were also made in terms of monetary policy targets with various changes applied to the statutory liquidity ratio and cash reserve ratio, as well as a slew of micro-prudential measures[2]. These myriad of policy changes has created a unique scenario for the Indian banking industry where the substantial part of the industry is dominated by public sector banks, with the national government as its majority stakeholder. The existing literature on the Indian banking in-

[2]Dharmarajan (2023) discusses these changes in greater detail

dustry can be separated into two main topics: (a) studies measuring efficiency of Indian banks and changes in efficiency over time; (b) studies measuring competition in the banking industry and its changes overtime.

Bhattacharyya et al. (1997) find public sector banks experienced lower productivity growth between 1970-1992 due to nationalization. Reviewing the period from 1985-1996 for both public and private sector banks, Kumbhakar and Sarkar (2003) they did not find evidence for any significant improvement in total factor productivity growth for both public and private sector banks following the reforms in the early 1990's. Casu et al. (2013) studied the post-reform period between 1992-2009 and found that there were differences in productivity gains by ownership. Foreign banks to have the most efficient production technology which was later adopted by public and private sector banks. Gulati and Kumar (2016) find similar evidence of foreign banks continuing to employ the most efficient production technology that is subsequently adopted by rest of the market using data from 2003-2013. Sengupta and Vardhan (2020), however, finds that efficiency gains have stagnated in this industry since 2011-12. They studied at 33 scheduled commercial banks between 2002-2018 and find increasing non-performing loans has reduced productivity growth in the banking sector.

The second strand of literature that focuses on measuring competition and changes in market structure over time include. Using Panzar-Rose H-statistic[3], Prasad and Ghosh (2007) find that Indian banking industry is characterized by monopolistic competition. They argue that the market being contestable where the incumbent firms are pricing close to competitive level because of threat of potential competition. However, they do not test whether the market is in long-run equilibrium which is necessary to validate the estimated results[4]. Perera et al. (2006) computes the Panzar-Rosse H-statistic for four countries in the Asian subcontinent - Bangladesh, India, Pakistan and Srilanka. For the time period, 1995-2003,

[3]Panzar-Rose H-statistic measures the sum of elasticities of the price of output to changes in price of inputs.

[4]The long-run equilibrium test for the market was introduced by Bikker and Haaf (2002)

they find monopolistic competition in the banking industry in Bangladesh, India, Pakistan and Srilanka. In India, they find that banks are efficiently moving away from labor intensive process to capital intensive ones as indicated by positive relationship between capital costs and bank revenue. They also find that in India and Srilanka, the traditional interest-based banking market is less competitive than the overall market. This could be because of the presence of foreign banks who have more fee and commission based products [5]. Sensarma (2008) uses a stochastic frontier analysis to measure profit efficiency of Indian banks for the period 1986-2005 and finds that profit efficiency and profit productivity declined over time despite the reforms in the 1990's. This is interpreted as a result of increased competition in the industry as increased competition is leading to lower profits. Zhao et al. (2010) uses a stochastic cost frontier model to estimate marginal cost of individual banks and then use the estimated marginal costs to test for persistence of profits in the Indian banking industry. They study the period 1992-2004, which is divided into pre and post periods (1992-1997 and 1998-2004). They find that the regulatory reforms in 1992 and 1998 has led to decreasing markup's in the market for loans in the post period (1998-2004) relative to the pre-period (1992-1997). That is, they find evidence of increasing competition in the market. Their estimates can be interpreted in a similar way as decreasing Lerner index for the market for loans. The line of reasoning in this strand of literature is that increases in efficiency is due to increasing competition as competitive pressures incentivise banks to undertake measures to improve their cost efficiency and profit efficiency.

The study that aligns closely with this paper is Das and Kumbhakar (2016). They use an input distance function approach to estimate marginal costs and markups for the period 1992-2012. They find that markups decreased between 1992-2002 while it increased in the latter period between 2002-2012. They further argue that large banks (measured in terms of their total assets) have the capacity to impose higher prices, especially on loans, and enjoy

[5]On the other hand, in Bangladesh and Pakistan, they find that the traditional interest-based banking market is more competitive than the overall market (which includes fee and commission based banking products).

significant market power. However, they don't incorporate multimarket contact in their analysis. Neither do they use a structural model to estimate market structure parameters. To the best of my knowledge, there are no studies on the Indian banking industry that either look at multimarket contact between banks or apply Bresnahan-Lau model, let alone both.

Under the multimarket contact hypothesis, firms that operate in the same markets, either in terms of products or geography or both, have less incentive to compete aggressively against each other if they fear retaliation from rivals in other markets. This was first put forward by Solomon (1970). I test whether this is the case in the market for Indian banks. We can view banks as firms selling a homogeneous product in various local markets. Brick and mortar banking dominates the Indian banking system, thus creating geographical overlaps in terms of bank's network of branches. Moreover, mergers and acquisitions as well as amalgamations among banks in the industry has increased the degree of multimarket contact over time. Banks in India require license to open new branches as well as close existing ones. In 2005, the Reserve Bank of India introduced a policy change in issuance of bank branch licenses. There were two key changes under this policy. First, rather than issue licenses for branches on a case-by-case basis, RBI switched to a system where banks submitted an Annual Branch Expansion Plan (ABEP) and RBI granted permission to open branches based on it. Second, the total number of branch licenses issued to a bank depended on the number of branches the bank was planning to open in designated "underbanked markets" [6]. The period of study chosen in this paper, 2005-2020, thus facilitates the study of multimarket contact hypothesis as multimarket contact becomes even more likely post this policy change. This in turn, increases the likelihood of tacit collusion between banks.

There is evidence from studies in other markets and countries that find evidence in support of the theory of mutual forbearance. Heggestad and Rhoades (1978) analyzed 187 major banking markets in the US and found that multimarket meetings between the dominant

[6]The markets are defined at the district level by RBI based on average number of persons per branch within a district relative to the national average. See Young (2021) for detailed description.

banks adversely affected the extent of rivalry within markets as measured by changes in market shares for deposits. Rhoades and Heggestad (1985) studied the same sample period but used profits and prices as the outcome variable and found partial evidence in support of the theory of mutual forbearance. Mester (1987), on the other hand, finds pro-competitive effect from multimarket links. The author studied the savings and loan industry in California and found that the interaction between market concentration and contact is important rather than the contact itself. The author found that consumers benefited more in markets with high concentration and high contact. Whalen (1996) is another study on the US banking industry where the author analyzed the deposit market of large, interstate US bank holding companies and found evidence in support of linked oligopoly theory. Pilloff (1999) find that multimarket contact among US banking organizations are significantly and positively related to their profitability. The author finds this effect to increase with multimarket contact. Haveman and Nonnemaker (2000) find that multimarket firms grow and enter the markets in which they have moderate multimarket contact. They also find that the impact of multimarket links is strongest in markets dominated by a few multimarket firms. Their study is also based on the savings and loans industry in California as Mester (1987). Similar evidence for growth and entry was found in the Spanish savings bank market by Fuentelsaz and Gomez (2006).

Coccorese and Pellecchia (2009) and Coccorese and Pellecchia (2013) find evidence in support of the theory of mutual forbearance in the banking industry in Italy. More recent evidence on the role of multimarket contact in the banking industry comes from Coble (2019). The author uses a discrete-choice model for deposit services in the US to estimate different supply functions at the market and bank levels to fit a partial collusion model. One of the findings in this study is that for collusion to exist, banks have to be competing in more than 164 markets (markets defined at the Metropolitan Statistical Area). The author finds that the US banking market is competitive and resembles the theoretical outcome under Nash-Bertrand equilibrium. There are studies on multimarket contact in other non-bank markets which find

evidence in support ,as well as against, the mutual forbearance theory. I am not describing or citing them as they are not relevant to the current study both in terms of methods and results.

The Bresnahan and Lau framework applied in this paper closely follows Coccorese and Pellecchia (2013) who have aggregated the demand and equilibrium condition to the market level to estimate parameters of the market structure. Coccorese and Pellecchia (2013) also incorporates multimarket contact into this framework. Coccorese (2009) applies the same framework to individual bank level data to determine market structure parameters in the Italian banking industry. However, limitations in terms of data restricts my analysis to the market level similar to Coccorese and Pellecchia (2013).

This study adds to the literature in two ways. First, this is the first paper to analyze the role of multimarket contact in the Indian banking industry. My results indicate that multimarket contact does impact a bank's income. I find a statistically significant and positive impact of multimarket contact on bank's earnings (measured as ratio of income to total assets). This finding is robust to model specification, different measures of multimarket contact as well as sample of banks (and markets). I use a reduced form fixed-effects model at the bank-level as well as a structural model by Bresnahan (1982) and Lau (1982) to study the multimarket contact hypothesis. Increasing multimarket contact and thereby, facilitating collusion among banks is an unintended consequence of the regulatory policy changes . Second, I find evidence of monopoly behavior among banks in India using a structural model. Existing studies use non-structural models to measure competition in the banking industry. As discussed earlier in this section, these non-structural measures have their limitations as they impose strong assumptions on the equilibrium conditions of the market. As discussed earlier, my findings add to findings from Das and Kumbhakar (2016). I also compute these alternative measures of competition upto the most recent period as benchmark against the structural measures.

The findings of this study raise policy implications and questions. Suppose current policies

for branch expansion are inadvertently facilitating collusion between banks through increasing multimarket contact. This has implications for the larger objective of increasing financial inclusion (both in terms of access to credit as well as avenue for savings) as banks might maintain higher markups in the credit market while offer lower interest rates in deposit markets fearing retaliation from rivals in other markets. Data at the district-branch-loan level is required to effectively test this hypothesis. However, such data is not available publicly. The next best method to evaluate the hypothesis is at higher levels of aggregation as I do in this study[7].

In the next section, I set out the different models used in my analysis. For the analysis related to multimarket contact, I construct both bank-level and market-level measures of multimarket contact. The latter is required for the structural model. In the latter sections, I discuss the data and the results from the analysis.

2.3 Measures of Competition

2.3.1 Multimarket Contact Models

Bank level model

We test the mutual forbearance theory at the bank level and estimate whether multimarket contact affects a bank's profitability. We estimate the following equation:

$$RoA_{it} = \alpha MMC_{it} + \beta MktVar_{it} + \gamma BankVar_{it} + \delta_i + \theta_t + \varepsilon_{it} \tag{2.1}$$

[7]I have run district-level analysis where possible to check for robustness of results.

where i = 1,...,N, t=1,...,T. N is the total number of banks while T is the total number of time periods. Here, RoA measures bank's profitability in each time period while MMC represents a measure of multimarket contact for each bank in each period. The parameter of interest is α. The measure of profitability, RoA, is expected to be directly related to the multimarket contact measure, i.e. we expect $\alpha > 0$ because if the theory of mutual forbearance holds, then banks with greater multimarket contact should enjoy greater profits. The variables used in this model are defined as:

- $RoA = \frac{\text{Operating Income}}{\text{Total Assets}}$: The variable measures bank's profitability each period. I am using operating income rather than profits so that the measure of profitability is not unduly biased by any subjective changes in balance sheet policies or due to inclusion of extraordinary or exceptional items.

- $MktVar$ represents a vector of market level variables that are weighed and redefined at the bank level. These variables are:

  - Herfindahl-Hirschmann Index (HHI) defined for each market based on the distribution of bank branches[8]. HHI is defined such that it ranges between 0 to 1. We expect markets with higher concentration to have higher likelihood of collusion under the SCP paradigm.

  - Total amount of credit issued in the market, with the variable defined as natural logarithm of all loans disbursed in the market. This measures the ease of entry as larger markets are easier to enter and therefore, more competitive. I expect a negative relationship between market size and bank's profitability.

  - Growth rate of credit in market, where the variable is defined as the growth in annual credit disbursed in each market. If the market for loans are expanding at

[8]Ideally, the market share should be defined either based on bank's deposits or loans in each market. However, due to non-availability of such data, the next best solution is to compute HHI using bank's market share in terms of branches. Other studies such as Coccorese and Pellecchia (2009), Coccorese and Pellecchia (2013) and Degryse and Ongena (2005) use the same methodology to compute market share of banks

a faster rate, then we would expect banks' profits to also increase over time as interest earned on credit is the primary source of income for banks.

  - The market level variables are weighted by bank's market share of branches in each state and then converted to bank-level variables[9].

- $BankVar$ are bank level variables computed off their annual financial statements. They include:

  - Degree of capacity utilization as measured by total assets per branch of the bank. We take the natural log of this variable. Greater capacity utilization is expected to have a positive impact on bank's profitability.

  - Potential for profits is measured by the ratio between bank's loans to its total assets. Higher the ratio, higher the potential gross yield, i.e. higher the potential interest income the bank can earn and therefore, we expect a positive effect on profitability.

  - Cost of funds to the bank is defined by the equity to assets ratio, where equity is the sum of capital and reserve surplus of a bank. A higher equity to asset ratio is represents a lower cost of funds, lower probability of bankruptcy and its expected costs, and therefore is expected to be positively associated with higher profits.

  - Ratio between non-interest operating expenses and total assets to represent bank's efficiency in its operations. A higher ratio reflects higher cost incidence and therefore, indicates lower profitability[10].

  - The natural logarithm of total number of bank's branches is also included to measure the effect of branch networks on profits. A wider network of branches allows bank to take advantage of more profitable opportunities as well as diversify their risks. On the flip side, a wider network of branches increases bank's costs.

[9] This is necessary as all bank-invariant variables will be removed by the fixed effects model.

[10] Interest payments are not included in this variable because they can be influenced by external macroeconomic factors and need not necessarily reflect the decisions of the bank's management.

Therefore, the impact of the size of a bank's network of branches on its profits can go either way.

  - As an indicator of potential profits, we also include the ratio of bad loans to total loans of a bank. Higher the ratio, higher the probability of losses (or less profits) for the bank.

- The multimarket contact measures, MMC_{it} is defined in greater detail in later subsections.

- δ_i and θ_t are bank and year fixed effects, respectively.

- ε_{it} is an error term with mean zero and finite variance.

Market level model

In brief, the following structural model takes advantage of the profit maximizing condition, i.e. MR = MC. We continue with the analysis of market for loans in the banking industry. We consider a set of banks, N, where each bank maximizes its profits by setting their output (i.e. credit disbursed) at the level where their marginal revenue equals their marginal costs. If the market is perfectly competitive, then marginal revenue will be equal to price, and we have P=MC as the profit maximizing condition. However, if the market is a monopoly or a market with perfect collusion, then the marginal revenue of each individual firm is equal to the marginal revenue for the entire market. We will use a simultaneous equations model, with a demand equation and an optimal equilibrium condition (MC=MR), where a parameter measuring average degree of market power of banks is included. This parameter is assumed to be function of multimarket contact and market concentration. Using the data on Indian banks and defining each state as a market, we seek to identify this parameter.

Consider the market demand for loan in region j at time t to be:

$$Q_{j,t} = Q_{j,t}(P_{j,t}, X_{j,t}, \delta)$$

where $Q_{j,t}$ is the aggregate level of loans in region j at time t, $P_{j,t}$ is the interest rate on loans charged by local banks in region j at time t, $X_{j,t}$ are vector of exogenous variables shifting the demand curve. δ is vector of unknown parameters to be estimated.

Aggregating individual demand curves and then taking the derivative with respect Q gives us the "true" MR curve for the entire industry, i.e. :

$$MR_{j,t} = P_{j,t} + \frac{Q_{j,t}(.)}{\frac{\delta Q_{j,t}}{\delta P_{j,t}}}$$

For each individual bank i, its perceived MR curve is:

$$MR_{i,j,t} = P_{j,t} + \lambda_{i,j,t}\frac{q_{i,j,t}(.)}{\frac{\delta Q_{j,t}}{\delta P_{j,t}}}$$

where $\lambda_{i,j,t}$ is an unknown parameter that measures competitiveness of oligopoly conduct. $\lambda_{i,j,t} = 0$ under perfect competition and $\lambda_{i,j,t} = 1$ under joint monopoly or perfect collusion. Note that if λ is negative, it could be that there are unsustainable deviations from the long-run equilibrium where banks are pricing below their marginal cost. In this case, the aggregate output $Q_{j,t}$ will exceed the competitive norm (Gruben and McComb (2003) and Shaffer (2004)).

We aggregate the individual demand and supply functions to solve the over-parametrization of the model. This also allows use market-level data rather than individual bank-market level data. In the context of this study, I define a market as a state[11].

[11]There are 35 different markets in India under this characterization.

Horizontal summation of individual banks' marginal costs functions gives us the industry marginal cost function, $MC_{j,t}$ Shaffer (2004). Moreover, this also means we now have only one parameter, $\lambda_{j,t}$, that measures the average conduct of banks operating in market j at time t (Bresnahan 1989). We can write the equilibrium condition MR=MC for the industry as:

$$MR_{j,t} = P_{j,t} + \lambda_{j,t} \frac{Q_{j,t}(.)}{\frac{\delta Q_{j,t}}{\delta P_{j,t}}} = MC_{j,t}$$

which can be re-written as:

$$P_{j,t} - MC_{j,t} = -\frac{\lambda_{j,t}}{\eta_{j,t}} \tag{2.2}$$

where $\eta_{j,t} = \frac{\frac{\delta Q_{j,t}}{\delta P_{j,t}}}{\frac{1}{Q_{j,t}}}$ is the semi-elasticity of market demand with respect to price. gives us a measure of deviation of price from marginal cost, i.e. the perfectly competitive price. Alternatively, Shaffer (1993) showed λ is a local approximation of the percentage deviation of total industry output from the optimal level of production that characterizes the perfectly competitive equilibrium.

We define this parameter, λ as:

$$\lambda_{j,t} = \lambda_{j,t}(MMC_{j,t}, HHI_{j,t})$$

where λ is defined as a function of multimarket contact and the concentration in the local market, as measured by MMC and HHI, respectively. Here, MMC is computed based on presence (or absence) of banks in each market and is described in more detail in the next section. HHI is the Herfindahl-Hirschman index which is defined to be between 0 and 1. The parameter λ is a function of local market concentration, HHI, to account for the possibility that banks may price above marginal cost in regional markets that are more concentrated.

As Shaffer (2004) points out, this approach provides us with a test statistic that maps into all oligopoly solution concepts and at the same time allows us to use data aggregated at the industry level in each market[12]. Shaffer (2001) also show that we do not have to expect any biased estimation of the market power parameter because we are including all potential markets rather than a subset of the industry. However, there is a reason why this method has not been applied frequently despite its advantages. The Bresnahan-Lau test requires a clear identification of the product or geographic market because it uses market level variables. The estimation is also done through a non-linear system of equations and requires data on aggregate demand.

The demand function for loans in each market j for year t is defined as follows:

$$lnQ_{jt} = \alpha_1 P_{jt} + \alpha_2 Z_t + \alpha_3 Y_{jt} + \alpha_4 POP_{jt} + \alpha_5 BR_{jt} + \alpha_6 t + \alpha_j \tag{2.3}$$

where Q_{jt} is the amount of loans in region j at time t, P_{jt} is the average market loan rate, Z_t is the interest rate on 10-year benchmark bond (as a proxy for substitutes for bank loans), Y_{jt} is the Net Domestic Product for the state j in period t. POP_{jt} is population in region j at time t, BR_{jt} is the number of branches in region j at time t, α_j represents fixed effects for regions to account for any unobserved regional characteristics. Here α_1 provides a measure of average value of price semi-elasticity of demand η. A time trend variable, t, is also included to capture possible trend effects. Here, we use a semi-logarithmic demand function because we do not have to impose constant elasticities, which can be an issue with time-series data.

The supply function for market j is derived from individual bank's supply function. For bank i in region j at time t, its supply function is defined in the form of a translog cost function with two generic inputs. One of the inputs is labor measured by number of employees for bank i in region j at time t. The other input is deposits which represents the funds required

[12]This approach also allows us to test for monopsony power (Shaffer (2004)) in the market for deposits. However, lack of bank level data on deposit rates for the entire duration of the sample prevents me from applying this model to the market for deposits.

to create loans for each bank. There is one output, q, which is the amount of loans disbursed. The supply function is defined as:

$$\begin{aligned} lnC_{ijt} = \beta_0 + \beta_Q lnq_{ijt} + \frac{\beta_{QQ}}{2}(lnq_{ijt})^2 + \beta_1 lnW_{1jt} + \beta_2 lnW_{2jt} + lnq_{ijt}(\beta_{Q1} lnW_{1jt} + \beta_{Q2} lnW_{2jt}) \\ + \frac{1}{2}[\beta_{11}(lnW_{1jt})^2 + \beta_{22}(lnW_{2jt})^2 + \beta_{12} lnW_{1jt} lnW_{2jt} + \beta_{21} lnW_{2jt} lnW_{1jt}] \\ + \beta_T lnTIME + \frac{\beta_{TT}}{2}(lnTIME)^2 + lnTIME(\beta_{QT} lnq_{ijt} + \beta_{1T} lnW_{1jt} + \beta_{2jt} lnW_{2jt}) \end{aligned} \tag{2.4}$$

where C_{ijt} and q_{ijt} are bank i's total cost and total output (loans), respectively, in region j for period t. TIME is the time trend variable (similar to demand equation) and it is included possible effects of technological change over time[13]. W_{1jt} and W_{2jt} are exogenous prices of inputs. W_{1jt} is the regional interest rate on deposits and W_{2jt} is the average regional wage rate of the financial intermediation sector. C_{jt} is the sum of interest expenses (computed as the product of deposit rate times total deposits) and labor costs (assumed to be equal to the salaries paid by banks). Both of these variables are defined at the regional level for each time period.

Similar to the discussion of estimation of marginal cost in the previous chapter on Lerner index and Boone indicator, we impose conditions of linear homogeneity in input prices on the cost function. We also assume symmetry, i.e. we have $\beta_{12} = \beta_{21}$. Assuming input prices are linear homogeneous implies: $\beta_1 + \beta_2 = 1$; $\beta_{11} + \beta_{12} = 0$; $\beta_{21} + \beta_{22} = 0$; $\beta_{Q1} + \beta_{Q2} = 0$; $\beta_{1T} + \beta_{2T} = 0$. To impose these conditions, we can divide both sides by W_{2jt} and then

[13]In the current specification, I have defined the time trend as TIME to avoid confusion with the subscript t

aggregate across banks in region j to get:

$$ln(\frac{C_{jt}}{W_{2jt}}) = \beta_0 + \beta_Q lnQ_{jt} + \frac{\beta_{QQ}}{2}(lnQ_{jt})^2 + \beta_1 ln\frac{W_{1jt}}{W_{2jt}}$$
$$+\beta_{Q1} lnQ_{jt} ln(\frac{W_{1jt}}{W_{2jt}}) + \frac{\beta_{11}}{2} ln(\frac{W_{1jt}}{W_{2jt}})^2 + \beta_T lnTIME +$$
$$\frac{\beta_{TT}}{2}(lnTIME)^2 + lnTIME[\beta_{QT} lnQ_{jt} + \beta_{1T} ln(\frac{W_{1jt}}{W_{2jt}})]$$

Taking the derivative with respect to the output, Q gives us the marginal cost function for the industry:

$$MC_{jt} = \frac{C_{jt}}{Q_{jt}}[\beta_Q + \beta_{QQ} lnQ_{jt} + \beta_{Q1} ln(\frac{W_{1jt}}{W_{2jt}}) + \beta_{QT} lnTIME] \tag{2.5}$$

Plugging 2.5 into 2.3.1 gives:

$$P_{jt} = \frac{C_{jt}}{Q_{jt}}[\beta_Q + \beta_{QQ} lnQ_{jt} + \beta_{Q1} ln(\frac{W_{1jt}}{W_{2jt}}) + \beta_{QT} lnTIME] - \frac{\lambda_{jt}}{\alpha_1}$$

We can also specify λ_{jt} as:

$$\lambda_{jt} = \lambda_0 + \lambda_1 M_{jt} + \lambda_2 HHI_{jt} + \lambda_3 M_{jt} * HHI_{jt}$$

Plugging this back into the equilibrium condition gives us:

$$P_{jt} = \frac{C_{jt}}{Q_{jt}}[\beta_Q + \beta_{QQ} lnQ_{jt} + \beta_{Q1} ln(\frac{W_{1jt}}{W_{2jt}}) + \beta_{QT} lnTIME] - \frac{\lambda_0 + \lambda_1 M_{jt} + \lambda_2 HHI_{jt} + \lambda_3 M_{jt} * HHI_{jt}}{\alpha_1} \tag{2.6}$$

2.3 and 2.6 have to be estimated simultaneously. Given the non-linearity in the functional specification, we need to apply non-linear 3SLS model for the estimation. To correctly identify the parameter, λ_{jt}, Lau (1982) show a necessary and sufficient condition in a system of demand and cost equations is that the demand function must not be separable in at

least one exogenous variable that is included in the demand function but excluded from the marginal cost function. The semi-logarithmic specification of the demand function allows us to meet this condition. For instance, $\frac{\delta^2 Q_{jt}}{\delta P_{jt} \delta POP_{jt}} = \alpha_1 \alpha_4 Q_{jt} \neq 0$, i.e. the two equations are non-separable with respect to population in the region (POP_{jt}).

The next section discusses the definition of how multimarket contact is defined and measured.

2.3.2 Multimarket contact measures

Bank-level Multimarket Contact measures

I construct multimarket contact measures at the level of the bank as well as the level of the market. The former is applied to the bank-level model while the latter is used in the market level structural model[14].

I will first define all the pieces required to construct the measures and then define the measures themselves. Consider a country with N banks and M markets. Let $BR = [br_{im}]$, where br_{im} is the number of branches of bank i in market m. We can now define another $N \times M$ matrix, $U = [u_{im}]$, where $u_{im} = 1$ if $br_{im} > 0$ and zero otherwise. Each element of the matrix U indicates whether bank i is present in market m (=1) or not (=0). Let $A = UU'$ where A is a NxN symmetric matrix with

$$a_{kl} = \Sigma_{m=1}^{M} u_{km} u_{lm}$$

Every off-diagonal element measures the number of markets in which bank k and l have contact and every diagonal element represents the number of markets in which bank k operates. We define another matrix $N \times M$, $S = [s_{im}]$ where each element, s_{im} is the market

[14]These mulitmarket measures are constructed based on Evans and Kessides (1994), Jans and Rosenbaum (1997), Coccorese and Pellecchia (2009), and Coccorese and Pellecchia (2013).

share of bank i in market m and -

$$s_{im} = \frac{br_{im}}{\Sigma_{i=1}^{N_m} br_{im}} * 100$$

We then construct a similarity matrix, $SI_{N \times N}$, which measures similarity between banks i and j. Similarity between the two banks i and j is defined as the product of absolute differences in their market shares in markets where they both meet.

$$SI_{ij} = \Sigma_{m=1}^{M} |s_{im} - s_{jm}| \times u_{im} \times u_{jm}$$

By construction, the similarity index ranges between 0 and the number of markets where the banks meet. The lower limit of the index is 0 when both banks have same market shares in the market in which they meet. The similarity index will be smaller if the banks have similar market shares.

I now define the three bank-level multimarket contact measures:

Measure 1: BMMC1

The first measure of mulitmarket contact at the bank level measures the average number of contacts for a bank, defined as:

$$BMMC1_i = \frac{\Sigma_{i \neq j} a_{ij} d_{ij}}{d_{ij}} \tag{2.7}$$

where $d_{ij} = 1$ if $a_{ij} > 0$ and 0 otherwise. The lower and upper limit of this measure depends on how the branches of the bank are distributed across the markets. The measure is 0 if a bank is a monopolist in every market it operates. On the other hand, if all banks meet each other in all markets, then the maximum value of this measure is reached which is equal to the number of markets. If an individual bank operates only in one market and is not a monopolist in that market, then BMMC1 = 1.

Measure 2: BMMC2

It is possible that banks may not treat every rival bank the same way. Banks might have greater incentive to collude with larger counterparts. To factor this in, I construct another measure of multimarket contact based on how similar two banks are. To construct this measure, I first create w_{ij} which rescales the similarity measure S_{ij} such that the rescaled variable is: $0 \leq w_{ij} \leq 1$, where,

$$w_{ij} = \frac{a_{ij} - SI_{ij}}{a_{ij}}$$

The second measure is defined as:

$$BMMC2_i = \frac{\Sigma_{j \neq i} a_{ij} \times w_{ij} \times \widehat{d_{ij}}}{\Sigma_{i \neq j} \widehat{d_{ij}}} \tag{2.8}$$

where $\widehat{d_{ij}} = 1$ if $a_{ij} \times w_{ij} > 0$ and 0 otherwise. By construction, $BMMC1_i \geq BMMC2_i$. They are equal only when bank i has same market share as all the rivals it meets with.

Measure 3: BMMC3

The third measure focuses on the size of the rivals rather than similarity with rivals. In a sense, this measure is an average of sums of rivals' market shares. It is defined as:

$$BMMC3_i = \frac{\Sigma_{j \neq i} a_{ij} \times \widetilde{w_{ij}} \times \widetilde{d_{ij}}}{\Sigma_{i \neq j} \widetilde{d_{ij}}} \tag{2.9}$$

where,

$$\widetilde{w_{ij}} = \frac{\Sigma_{m=1}^{M} s_{jm} \times u_{im} \times u_{jm}}{a_{ij}}$$

and $\widetilde{d_{ij}} = 1$ if $a_{ij} \times \widetilde{w_{ij}} > 0$ and 0 otherwise.

Market-level Multimarket Contact measures

We continue to use the matrices $BR_{N\times M}$, $U_{N\times M}$, $A_{N\times N}$, and $S_{N\times M}$ from the previous section. To construct market level multimarket contact measures, we construct additional matrices at the market level.

Let $B^{(-m)}$ be a $N \times N$ symmetric matrix with each element, $b_{kl}^{(-m)}$ is the sum of market shares of banks k and l in all markets other than m, i.e.

$$b_{kl}^{(-m)} = \Sigma_{(p\neq m)}(s_{kp} + s_{lp}) * u_{kp}u_{lp}$$

We can also define another $N \times N$ symmetric matrix, $C^{(-m)}$, where each element is given by the concentration indexes characterizing the "non-home" markets (i.e. markets other than j) in which banks k and l meet:

$$c_{kl}^{(-m)} = \Sigma_{(p\neq m)} HHI_p * u_{kp} * u_{lp}$$

with

$$HHI_m = \Sigma_{i=1}^{N_m} ms_{im}^2$$

Measure 1: MMC1

Our first measure of multimarket contact measures extent to which two banks meet in regions other than m for every contact made in market m. Hence, we define $MMC1_m$ for market j:

$$MMC1_m = \frac{\Sigma_{k=1}^{N-1}\Sigma_{l=k+1}^{N} a_{kl}u_{km}u_{lm} - \frac{N_m(N_m-1)}{2}}{\frac{N_m(N_m-1)}{2}} \tag{2.10}$$

where N_m is the number of banks in market j and $\frac{N_m(N_m-1)}{2}$is the total number of possible

pairings between any two banks. The measure, $MMC1_m$ does not factor the relative size of these banks in each market, i.e, it fails to account for the concentration within each market j. Collusion is more likely in highly concentrated markets as the costs of retaliation from competitors are larger and the costs of coordinating is smaller with fewer players in the market. Therefore, we can augment $MMC1_m$ with the market shares of each bank.

Measure 2: MMC2 The augmented measure of multimarket contact is:

$$MMC2_m = \frac{\Sigma_{k=1}^{N-1}\Sigma_{l=k+1}^{N} b_{kl}^{(-m)} u_{km} u_{lm}}{\Sigma_{k=1}^{N-1}\Sigma_{l=k+1}^{N} a_{kl} u_{km} u_{lm} - \frac{N_m(N_m-1)}{2}} \tag{2.11}$$

This measures the average sum of market shares in markets other than j for banks operating in market j per contact between banks in market j.

Measure 3: MMC3 A third measure of multimarket contact is defined as:

$$MMC3_m = \frac{\Sigma_{k=1}^{N-1}\Sigma_{l=k+1}^{N} c_{kl}^{(-m)} u_{km} u_{lm}}{\Sigma_{k=1}^{N-1}\Sigma_{l=k+1}^{N} a_{kl} u_{km} u_{lm} - \frac{N_m(N_m-1)}{2}} \tag{2.12}$$

Here, we consider the regional vlaues of HHI to compute MMC3 as the average HHI of non-home markets per contact in the same markets.

We use each of these measures in 2.6 to estimate λ.

2.4 Data Sources and Summary Statistics

The data used for this analysis is compiled from multiple sources. The two primary sources of data are from the Database on Indian Economy (DBIE), published by the Reserve Bank of India and annual "Performance Highlights" reports published by the Indian Banks' Association (IBA). The latter was used to complement and fill gaps in data from the former source when necessary. The data is published according to the Indian financial calendar

where the financial year spans from the month April of a given year to March of the next year.

Data can be classified into the following categories based on their source level of disaggregation:

- Data from DBIE (RBI):
  - Annual financial statements of scheduled commercial banks for the period 2003-2021. This data is at the country-bank-year level.
  - Basic Statistical Returns of Scheduled Commercial Banks to compile data on credit and deposits. This data is at state-bank group-year level.
  - Bank Branch Statistics covering data on number of branches and number of new offices opened. The data is at the bank-district/state-quarter level covering 2005-2021.
- Data from IBA:
  - Annual performance highlights, which includes financial statements between 2003-2020. The data is at the country-bank-year level.
- Data from Ministry of Statistics and Program Implementation (MOSPI):
  - Index of Industrial Production - aggregate index and its components at a monthly frequency.
  - Consumer Price Index and Wholesale Price Index - aggregate index data at a monthly frequency.
- Data on interest rates charged by Public, Private and Foreign banks were collected from RBI's website through web-scraping.

Two other datasets used for this study is data on GDP Deflator which was downloaded from FRED database. I collected data on the daily yield on the 10-year benchmark bond from Bloomberg.

We have data on unique 141 scheduled commercial banks during the sample period spanning from March, 2004 to March, 2021 [15] . Though certain data is available at the district-level and on a quarterly frequency, I have restricted the analysis to annual frequency and defined markets at the level of a state[16]. This is because all the variables used in the analysis are available at an annual frequency and at the state level. That is, data is not available for all variables at the district-level and on a quarterly basis to perform analysis at the disaggregated level and frequency. Additionally, banks have merged with each other or certain public sector banks have been amalgamated into larger public sector banks during the sample period. I have considered each bank as separate entity until the date of their merger/amalgamation. There are also two instances were a state was split into multiple parts during the sample period [17] As data is not available at disaggregated level for these newly created states or administrative regions, I have defined them to be a single market for the duration of the analysis.

For the bank-level analysis, we consider all banks - banks that operate in multiple markets as well as single market banks - the analysis[18]. The final dataset consists of 110 unique banks spanning the years 2005-2021 over 34 different markets (states). The panel is not balanced as some banks exit the industry or merge or be amalgamated with other banks during the

[15]Even though financial data is available for years prior to 2005, data on other variables are not available and cannot be imputed from available data. For instance, data on branches by state (or district) for each bank is available only starting March, 2005. However, I am able to compute the same data for 2004 using related data on branch openings that covers the financial year 2004.

[16]States represent the higher level of administration above districts

[17]A new state of Telegana was created out of the existing state of Andhra Pradesh in 2014. Similarly, the state of Jammu and Kashmir was split into two union territories - Jammu and Kashmir, and Ladhakh.

[18]As robustness check, we consider the subset of public and private banks only for the analysis. The results are reported in Appendix A and they indicate evidence similar to the results with all banks included. I could not run separate models for each bank group because of lack of degrees of freedom given the number of fixed effects and parameters to be estimated.

sample period. All amounts have been deflated using GDP Deflator and are defined in terms of 2011-12 Indian Rupee (INR).

Tables 1 and 2 list the summary statistics of the key variables used in the bank-level model used to test the mutual forbearance hypothesis in the presence of multimarket contact.

Table 2.1: Bank-wise Summary statistics for Bank-level model of Multimarket Contact

Variable Name	N	Mean	Standard Deviation	Median	Min	Max
BMMC 1	1386	8.29	5.61	9.73	1.00	15.58
BMMC 2	1386	8.03	5.41	9.52	0.77	15.36
BMMC 3	1386	0.18	0.13	0.19	0.01	0.48
Total Earnings (INR Million)	1386	91935.12	196821.88	25690.93	61.44	2295241.25
Total Assets (INR Billion)	1386	1102.91	2467.63	306.69	0.43	30964.99
Return on Assets (%)	1386	8.60	2.80	8.35	0.47	35.81
Bad Loans (INR Million)	1386	17569.92	51669.08	1279.63	0.00	873606.31
Total Loans (INR Million)	1386	642330.50	1470431.88	155704.22	1.82	17640678.00
Equity (INR Billion)	1386	83.23	174.91	29.73	0.07	1760.11
Total Number of Branches	1386	1159.87	2342.07	246.00	1.00	24949.00
Operating Expenses (INR Million)	1386	19509.42	44895.88	5825.86	29.81	570301.00

Table 2.1 reports the summary statistics of the variables used in the bank-level model for multimarket contact. These summary statistics have been averaged across all banks and across all years. The mean and the standard deviation of these variables indicate a lot of variation across banks and across time. BMMC1 and BMMC2 have a mean of 8.29 and 8.03 respectively. As mentioned in section 3, the key difference between BMMC1 and BMMC2 is that the latter is weighted by how similar the two banks (in terms of market share for branches) in the markets in which they meet. The third multimarket contact measure (MMC), BMMC3 has a mean of 0.18. By construction this measure is takes values

between 0 and 1. All rupee variables have been deflated using the GDP Deflator and are measured in 2011-12 Indian Rupee (INR). I have also scaled the variables into billions or millions depending on the distribution of its values. Total number of branches is represented in actual units. All the statistics reported are for public sector banks, private sector banks, and foreign banks.

Table 2.2 describes the summary statistics of all state level variables used in the bank-level analysis of multimarket contact. Similar to the variables in Table 2.1, all rupee denominated variables have been deflated and converted to 2011-12 INR using the GDP deflator. To apply the bank-level analysis, each of these variables have been converted to level of the bank by taking weighted sum of each of these state-level variable for every bank in the sample. The weights are defined as the market share of the given bank in the given year in each state, where the market share are computed using the branches [19].

Table 2.3 reports the market level summary statistics of variables used for the market-level Bresnahan-Lau model. Similar to the bank-level variables, I have deflated each of the rupee denominated variable by the GDP deflator to define all these variables in terms of 2011-12 INR. The average interest rate on loans during the sample period is 10% per annum (after rounding)[20].

Figure 2.1 provides a visual representation of the average multimarket contact measure (constructed at the market level) for sample period across states. I have divided the states into quintiles on the basis of these multimarket contact measures. On average (across the sample period), we can see variation in the three market level measures of multimarket contact.

[19] I do not have access to any other bank-wise state-level variable such as loans, deposits, assets etc that can be used instead.

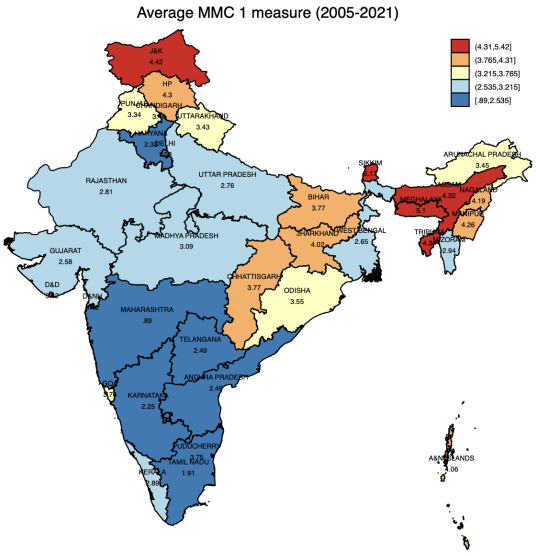
[20] This is not the imputed interest rate on loans as it can be biased by the presence of non-performing loans which became a severe issue during the latter half of the sample period.

Table 2.2: State-wise Summary statistics for Bank-level model of Multimarket Contact

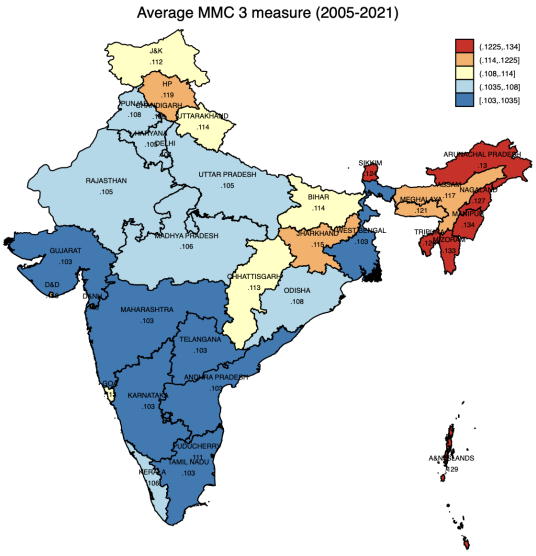
Variable	N	Mean	Standard Deviation	Median	Min	Max
State HHI	577	0.16	0.14	0.10	0.04	0.82
State Population (millions)	577	38.50	45.60	25.53	0.00	231.14
Total Branches in State	577	2786.82	3178.32	1503.00	10.00	13794.00
Total Banks in State	577	38.06	16.38	38.00	2.00	82.00
Total Number of Credit Accounts (millions)	577	3.78	6.65	1.14	0.00	54.94
Growth rate in Credit Accounts (%)	577	10.92	15.00	8.14	-35.71	143.33
Total Amount of All Credit (INR Billion)	577	1468.38	2989.39	331.35	0.23	21604.10
Growth rate in All Credit (%)	577	11.60	11.66	10.07	-20.81	104.62
Total Number of Small Borrower Credit Accounts (millions)	577	2.95	5.35	0.85	0.00	42.54
Growth in Number of Small Borrower Credit Accounts (%)	577	9.55	20.47	5.86	-50.00	178.62
Total Credit to Small Borrowers (INR Billion)	577	121.49	187.90	39.76	0.13	1066.22
Growth in Credit to Small Borrowers (%)	577	4.99	16.26	2.96	-59.06	100.01
Total Number of Deposit Accounts (millions)	577	30.79	41.97	13.11	0.03	217.36
Growth in Number of Deposit Accounts (%)	577	9.90	7.55	9.77	-18.16	52.51
Amount of Total Deposits (INR Billion)	577	1930.93	3168.51	697.74	2.02	21379.99
Growth in Total Deposits (%)	577	9.33	7.84	8.82	-19.48	50.66
Total Number of Current Deposit Accounts (millions)	577	1.36	1.93	0.60	0.00	14.47
Growth in Number of Current Deposit Accounts (%)	577	13.30	26.12	9.38	-74.85	199.91
Total Amount of Current Deposits (INR Billion)	577	203.98	407.72	69.45	0.08	3233.40
Growth in Current Deposits (%)	577	10.10	28.89	6.74	-72.23	264.79
Total Number of Savings Deposit Accounts (millions)	577	24.35	34.51	9.98	0.03	184.74
Growth in Number of Savings Deposit Accounts (%)	577	11.24	8.59	10.85	-25.22	77.57
Total Amount of Savings Deposits (INR Billion)	577	549.68	730.46	262.97	1.19	4912.68
Growth in Savings Deposits (%)	577	549.68	730.46	262.97	1.19	4912.68
Total Number of Term Deposit Accounts (millions)	577	5.08	6.15	2.62	0.00	30.75
Growth in Number of Term Deposit Accounts (%)	577	4.99	10.35	5.31	-51.62	89.52
Total Amount of Term Deposits (INR Billion)	577	1177.27	2142.18	372.51	0.63	14071.91
Growth in Term Deposits (%)	577	8.94	13.69	8.18	-44.99	177.10
Per capita number of deposit accounts in state (millions)	577	0.98	0.74	0.79	0.10	3.87
Per capita number of credit accounts in state (millions)	577	0.10	0.08	0.07	0.02	0.48
Per capita number of small borrower credit accounts in state (millions)	577	0.08	0.06	0.05	0.01	0.39
Per capita number of current deposit accounts in state (millions)	577	0.05	0.05	0.03	0.00	0.27
Per capita number of savings deposit accounts in state (millions)	577	0.74	0.51	0.63	0.08	2.58
Per capita number of term deposit accounts in state (millions)	577	0.19	0.22	0.13	0.01	1.30
Per capita amount of deposits in state (INR Thousands)	577	76.18	99.65	43.36	6.18	478.37
Per capita amount of credit in state (INR Thousands)	577	51.76	96.80	20.42	1.66	516.18
Per capita amount of credit to small borrowers in state (INR Thousands)	577	3.52	2.68	2.52	0.84	15.01
Per capita amount of current deposit in state (INR Thousands)	577	7.60	10.43	4.31	0.69	79.25
Per capita amount of savings deposit in state (INR Thousands)	577	21.78	22.08	14.98	2.29	145.83
Per capita amount of term deposit in state (INR Thousands)	577	47.03	69.85	22.42	1.94	341.69

Table 2.3: Summary statistics of variable used in market-level multimarket contact models

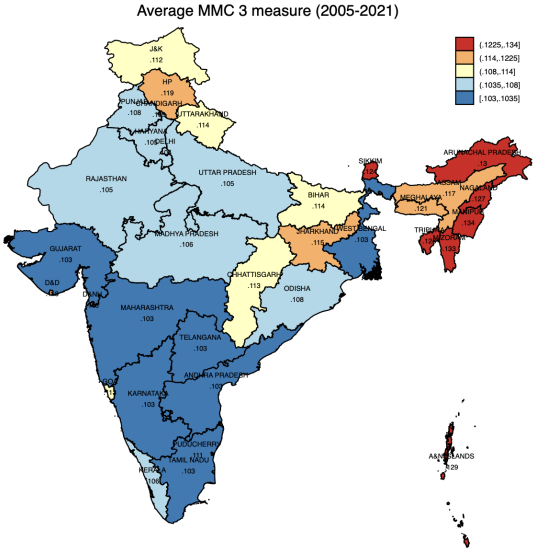
Variable	N	Mean	Standard Deviation	Median	Min	Max
MMC1	465	3	6	2	-1	32
MMC2	465	6	2	5	4	16
MMC3	465	0	0	0	0	0
Change in loans per period (INR Billion)	465	125	258	31	-311	1929
P (interest rate on loans) (%)	465	10	2	10	8	14
Yield on 10-year government benchmark bond (%)	465	8	1	8	6	9
Net State Domestic Product (INR Million)	465	2738730	3185048	1676287	35601	17829028
State Population (Millions)	465	40	46	27	1	228
Total Number of Branches in State	465	3055	3189	1951	28	13649
Average Variable Costs (INR Billions)	465	6	12	1	0	62
w1:Interest on Deposits (%)	465	5	1	5	4	7
w2:Wage per employee (INR thousands)	465	74	13	75	47	116
Total deposits in State (INR Billion)	465	2116	3241	881	12	21099
State HHI (based on branches)	465	0	0	0	0	1



(a) State-wise average MMC measure 1 (2005-2021)



(b) State-wise average MMC measure 2 (2005-2021)

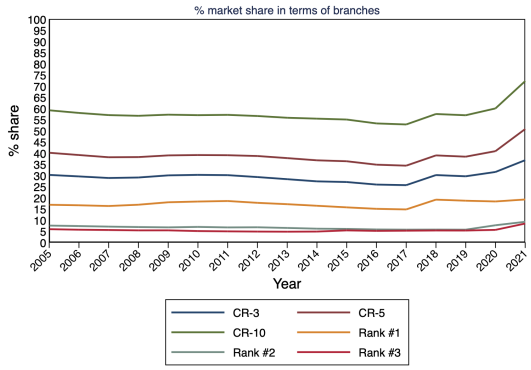


(c) State-wise average MMC measure 3 (2005-2021)

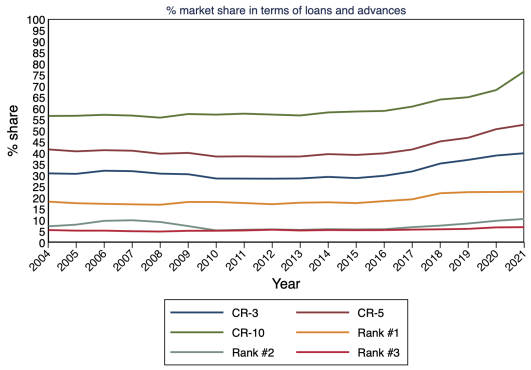
Figure 2.1: Statewise distribution of market-level MMC measures (average over 2005-2021)

For the set of public sector, private sector and foreign banks, Figures 2.2, plots the 3-firm, 5-firm and 10-firm concentration ratios each year in terms of number of active branches, loans and advances, deposits, and total assets of these banks. All four figures indicate simi-

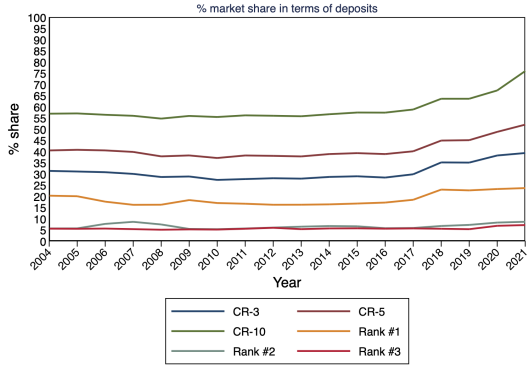
lar evidence - there is a dominant firm in the market which has a market share nearly double that of the next largest bank in the industry. The dominant bank discussed here is the State Bank of India, a public sector bank, and it maintains its dominance in market share throughout the entire sample period. Moreover, we observe an increase in concentration ratio towards the end of our sample period.



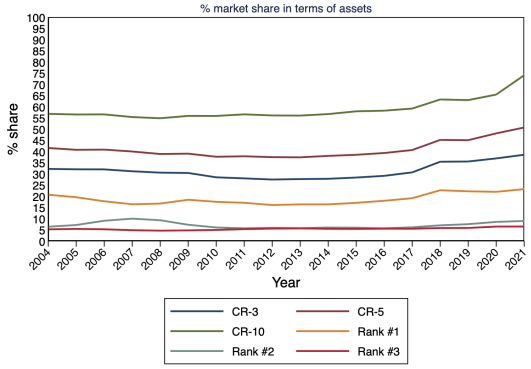
(a) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of number of branches, 2005-2021



(b) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of loans and advances, 2004-2021



(c) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of deposits, 2004-2021



(d) CR-3, CR-5, CR-10, Market shares of top 3 ranked banks in terms of total assets, 2004-2021

Figure 2.2: Market shares and concentration ratios - 2004-2021

Before we discuss the results of the various models, I want to highlight a few patterns based off the summary statistics. From Figures 2.2, we can see that concentration ratios have

increased since 2012. This is the case irrespective of whether we compute market shares in terms of branches or outstanding loans or deposits or total assets of banks. Moreover, we can also see that the top bank's share is increasing towards the end of the sample. The bank with the highest market share throughout the whole sample period is State Bank of India (SBI), a public sector bank. It's market share is increasing towards the end of the sample period as a few other smaller public sector banks have been amalgamated with it.

2.5 Results

Table 2.4[21] reports the estimates of running with different specifications where the multimarket contact is measured using BMMC-1[22]. The specifications differ in terms of the variables included. In all six specifications, I have included bank-level variables as well as time and bank fixed effects. In model (1), the only market level variable, i.e. state level variable, included is the HHI, as measured by share of bank's branches in the state. For each year, the state HHI measure was converted to bank level depending on bank's branch network each year. In the Model (2), I have included state-level variables[23] in level terms. The state level variables that are include are number of all credit accounts, number of small borrower credit accounts, number of current deposit accounts, number of savings deposit accounts, number of term deposit accounts, as well as total amount in each of these account types. In model (3), I have included both the first set of variables as well as the growth rate in the number of each of these accounts and the amounts in these accounts. In model (4), I have also included natural logs of the levels of the first of variables along with their growth rates. Model 5

[21] All specification and results reported use fixed effects model as Hausman test rejects the random effects model

[22] Model results using MMC measures defined at the district level are included in Tables B.4, B.5 and B.6 in Appendix B. The results reported in the main tables here are qualitatively similar to the results from using district-level MMC measures.

[23] These variables we converted to bank level variables based on distribution of a bank's branch network across states each year. This is necessary to avoid dropping these state level variables in to the presence of both time and bank fixed effects.

includes per capita values of each of these variables (in level terms) where I converted the level terms into per-capita terms using the state population each year. Model 6 includes the same set of state-level variables as Model 4 but I have added the state population as our objective is to factor for potential market size. The same pattern applies in Tables 5 and 6, where the only difference is in the multimarket contact measure used in the estimation.

Table 2.4: MMC 1 - Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 1	0.224***	0.223**	0.205***	0.245***	0.184***	0.250***
	(0.0670)	(0.0859)	(0.0744)	(0.0771)	(0.0573)	(0.0789)
NPA/Total Loans	-2.280	-1.730	-1.851	-1.970	-0.376	-2.058
	(2.533)	(2.452)	(2.501)	(2.589)	(1.923)	(2.597)
Total Assets per Branch	0.000232	0.00498	0.00425	0.00198	0.00381	0.00217
	(0.00662)	(0.00855)	(0.00891)	(0.00853)	(0.00901)	(0.00817)
Total Loans/Total Assets	0.132	-0.0295	-0.00704	0.0679	0.481	0.0643
	(1.428)	(1.291)	(1.283)	(1.270)	(0.885)	(1.256)
Equity/Total Assets	2.448	2.222	2.109	2.054	2.400	2.098
	(1.739)	(1.593)	(1.610)	(1.607)	(1.681)	(1.593)
ln(Branches)	0.534**	0.501***	0.523***	0.526**	0.556**	0.479**
	(0.239)	(0.164)	(0.165)	(0.202)	(0.215)	(0.229)
Operating Expenses/Total Assets	62.73***	63.58***	63.98***	62.12***	64.10***	60.70***
	(12.24)	(11.40)	(11.19)	(11.85)	(11.75)	(11.68)
HHI	-21.48**	-21.62**	-19.97	-12.85	-22.09**	-24.44*
	(10.83)	(9.653)	(12.08)	(13.22)	(10.53)	(12.92)
Observations	1386	1386	1386	1386	1386	1386
Within R-Squared	0.282	0.311	0.317	0.318	0.332	0.320
Between R-Squared	0.254	0.285	0.287	0.257	0.266	0.243
Overall R-Squared	0.212	0.237	0.239	0.217	0.231	0.208
Adjusted R-Squared	0.269	0.294	0.294	0.295	0.315	0.297
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

The results in Table 2.4 show strong evidence that multimarket contact positively and significantly impact's a bank's earnings as measured by its RoA. RoA is defined as the ratio of the bank's total earnings to its total assets and is measured in percentage terms. From Table 1, we know the median RoA for a bank during the sample period is 8.35% while the mean RoA is 8.60%. There a 1 unit increase in the multimarket measure is associated with an increase in a bank's RoA by 18.4 to 25 basis points, on average[24].

I also find that state level HHI negatively affects bank's RoA, i.e. increasing concentration in markets reduces the earnings of banks. State level HHI measure has a negative coefficient across all 6 specifications but is statistically significant at 5% level only in specifications (1), (2) and (5). The negative coefficient contradicts the SCP paradigm that states that firms in more concentrated markets should be earning higher profits. I interpret this as evidence in support of the mulitmarket contact hypothesis where competition can also be intense in markets with higher concentration ratios because it is the number of linkages across markets (as measured by the multimarket contact) that affects a bank's earnings [25]. Also, the HHI measure is constructed using market shares measured in terms of branches of banks and not the output. I do not have data on bank-wise distribution of credit (the output variable) in each market which means HHI can only be computed using the distribution and share of branches in each market. A higher HHI can also represent an inefficiently large branch network which negatively affects earnings. For instance, too many branches in the same market could lead to self-cannibalization of business across branches of the same bank[26].

The positive and significant coefficients for the ratio of operating expenses to total assets

[24]I have interpreted the result in terms of basis points as it we are measuring RoA in percentage terms. I have also taken the minimum and maximum estimates for the multimarket measure across all 6 specifications from Table 4 here.

[25]I also include the interaction terms between the multimarket measure and HHI for the same set of specifications. The interaction terms is never significant (at 5% level) in any of the specifications. Moreover, the absolute magnitude of the coefficient for HHI decreases while the magnitude of the coefficient for multimarket measure increases with the inclusion of the interactive term

[26]We can address this with a more granular definition of the market to address this issue but unavailability of data restricts this exercise.

adds to the evidence that it is not necessarily the more efficient banks that earn higher profits. Therefore, multimarket linkages allows inefficient banks to earn higher profits, consistent with the mutual forbearance theory[27].

Table 2.5: MMC 2 - Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 2	0.223***	0.224**	0.207***	0.247***	0.183***	0.252***
	(0.0685)	(0.0893)	(0.0781)	(0.0801)	(0.0581)	(0.0818)
NPA/Total Loans	-2.302	-1.743	-1.864	-1.995	-0.388	-2.084
	(2.533)	(2.450)	(2.499)	(2.586)	(1.921)	(2.595)
Total Assets per Branch	0.000260	0.00498	0.00424	0.00201	0.00382	0.00220
	(0.00664)	(0.00858)	(0.00894)	(0.00856)	(0.00904)	(0.00819)
Total Loans/Total Assets	0.136	-0.0246	-0.00219	0.0715	0.483	0.0679
	(1.427)	(1.289)	(1.281)	(1.268)	(0.885)	(1.254)
Equity/Total Assets	2.455	2.236	2.122	2.065	2.407	2.109
	(1.741)	(1.595)	(1.612)	(1.610)	(1.683)	(1.596)
ln(Branches)	0.549**	0.514***	0.535***	0.538***	0.569***	0.492**
	(0.238)	(0.163)	(0.164)	(0.200)	(0.213)	(0.227)
Operating Expenses/Total Assets	62.69***	63.52***	63.93***	62.06***	64.05***	60.63***
	(12.24)	(11.39)	(11.19)	(11.86)	(11.75)	(11.69)
HHI	-20.66*	-21.02**	-19.46	-12.11	-21.46**	-23.67*
	(10.80)	(9.641)	(12.06)	(13.20)	(10.47)	(12.84)
Observations	1386	1386	1386	1386	1386	1386
Within R-Squared	0.281	0.311	0.317	0.318	0.331	0.320
Between R-Squared	0.255	0.286	0.288	0.257	0.267	0.244
Overall R-Squared	0.213	0.238	0.240	0.217	0.231	0.209
Adjusted R-Squared	0.269	0.294	0.294	0.296	0.315	0.297
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

[27]An alternative explanation could be that larger banks with higher operating costs to total assets ratio maybe producing a higher quality of output which allows them to earn to greater revenues at the same time. This requires further analysis.

The other bank level coefficients, except for the log of branches and the ratio of operating expenses to total assets, are not statistically significant across different specifications. The positive and significant coefficient for the natural log of total number of branches for a bank indicates that banks with a wider network are able to earn more earnings. This is consistent with previous study (Prasad and Ghosh 2005) that find branch rationalization is one of the reasons for improvement in performance of Indian banks. In the context of multimarket contact, we can also interpret this as evidence in support of mutual forbearance theory. That is, banks with larger branch networks are more likely to have higher multimarket contact with rivals and thus greater opportunities for collusion[28].

The results in Table 2.5[29] where I use the second bank-level multimarket measure is almost identical to the results in Table 2.4 in terms of magnitude and sign of the coefficient as well as the significance. Therefore, we can conclude that even after we define a multimarket contact measure after taking into account similarity with rivals (where the number of contacts between two banks is weighted by an index measuring their similarity in terms of market shares in every market). The positive and significant coefficients for the second bank-level MMC measure is consistent with industrial organization theory that symmetry among firms can facilitate collusion (Barla 2000, Compte et al 2002).

In Table 2.6[30], the coefficient for the third multimarket contact measure is positive but not always statistically significant. The magnitude of the coefficient is also much larger but that is by construction as this measure is defined to be between 0 and 1. The interpretation of the positive coefficient, however, is consistent with the mutual forbearance hypothesis. For a given bank, larger the size of the rival's market share in which the two banks meet,

[28]This is further supported by lack of economic and statistical significance of the coefficient for total assets per branch across all specifications. This coefficient represents a bank's ability to utilize its capacity generate income.

[29]All specification and results reported use fixed effects model as Hausman test rejects the random effects model

[30]All specification and results reported use fixed effects model as Hausman test rejects the random effects model

greater the multimarket contact measure, and therefore, greater the earnings of the given bank (represented by higher RoA). As market shares have to sum to 1, this means that despite having smaller market shares relative to its rivals, banks can earn higher profits. This implies banks are colluding with their larger rivals, consistent with the predictions of the mutual forbearance hypothesis.

Table 2.6: MMC 3 - Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 3	4.459**	3.492*	3.245*	3.490	4.284***	3.472
	(1.877)	(1.827)	(1.951)	(2.212)	(1.618)	(2.197)
NPA/Total Loans	-2.427	-1.831	-1.949	-2.043	-0.406	-2.129
	(2.637)	(2.580)	(2.620)	(2.735)	(1.985)	(2.747)
Total Assets per Branch	0.000116	0.00520	0.00442	0.00208	0.00401	0.00225
	(0.00625)	(0.00831)	(0.00873)	(0.00834)	(0.00873)	(0.00801)
Total Loans/Total Assets	0.111	-0.0507	-0.0229	0.0643	0.492	0.0600
	(1.462)	(1.348)	(1.331)	(1.311)	(0.903)	(1.299)
Equity/Total Assets	2.208	2.040	1.931	1.834	2.216	1.871
	(1.729)	(1.584)	(1.599)	(1.599)	(1.669)	(1.587)
ln(Branches)	0.686***	0.675***	0.682***	0.726***	0.659***	0.690***
	(0.230)	(0.182)	(0.192)	(0.214)	(0.221)	(0.233)
Operating Expenses/Total Assets	62.41***	63.36***	63.76***	61.72***	64.17***	60.38***
	(12.15)	(11.21)	(11.01)	(11.63)	(11.71)	(11.50)
HHI	-18.97	-18.25*	-17.13	-11.41	-21.60**	-22.00*
	(11.88)	(10.71)	(11.13)	(12.63)	(10.73)	(11.44)
Observations	1386	1386	1386	1386	1386	1386
Within R-Squared	0.279	0.307	0.314	0.313	0.331	0.315
Between R-Squared	0.292	0.321	0.319	0.295	0.296	0.280
Overall R-Squared	0.249	0.273	0.271	0.255	0.261	0.246
Adjusted R-Squared	0.266	0.290	0.291	0.291	0.314	0.292
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

The coefficient for state HHI measure remains negative but is no longer significant across all specifications (at 5% level). The sign and significance of the other bank-level coefficients remains unchanged and follows the same interpretation as in the previous two tables. The analysis presented so far implicitly assumes a linear relationship between the multimarket contact measure and bank's profitability[31].

The next set of tables report the results from simultaneously estimating the system of equations 2.3 and 2.6 using non-linear 3SLS models. All exogenous variables in 2.3 and 2.6 are used as instruments (added as levels and logs). Moreover, because Q_{jt} and P_{jt}, I have included their first-lags as instruments to address any potential correlation between these variables and the error term. I have also included additional instruments in the form total value regional deposits (level and logs).

Table 2.7: Results of market-level structural model - MMC1

Variable		Without MMC - Model 1		MMC_1 - Model 2		MMC_1 - Model 3		MMC_1 - Model 4	
		Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
P	a_1	-.368***	(.062)	-.359***	(.047)	-.242***	(.048)	-.112**	(.044)
10-yr govt bond yield	a_2	1.095***	(.1)	1.235***	(.076)	.986***	(.078)	.918***	(.07)
NSDP	a_3	0	(0)	0	(0)	0	(0)	0	(0)
State Population	a_4	-.012***	(.003)	-.01***	(.002)	-.007***	(.002)	-.019***	(.002)
Branches	a_5	0**	(0)	0***	(0)	0	(0)	0***	(0)
T(time)	a_6	.282***	(.022)	.154***	(.017)	.25***	(.018)	.15***	(.015)
Marginal Cost Constant	b_Q	1.074	(99.876)	1.071	(93.354)	1.14	(97.787)	1.158	(210.718)
ln(Q)	b_{QQ}	.058	(13.361)	-.016	(12.489)	.045	(13.082)	.167	(28.193)
$ln(\frac{w1}{w2})$	b_{Q1}	.534***	(.073)	.303***	(.079)	.556***	(.075)	.755***	(.167)
ln T	b_{QT}	-.052***	(.018)	-.081***	(.017)	-.01	(.019)	-.17***	(.048)
Lambda Constant	λ_0	3.773***	(.643)	3.806***	(.507)	2.475***	(.502)	1.484**	(.58)
MMC1	λ_1			-.036***	(.009)	-.007*	(.004)	-.097**	(.041)
HHI	λ_2					.28	(.345)	-2.982**	(1.304)
MMC1*HHI	λ_3							.761**	(.319)
Observations		465	465	465	465	465	465	465	465

In Table 2.7, the first model does not include any MMC measure because we want to identify the structural parameter λ for the market. In the second model, I add the multimarket

[31]Existence of non-linear relationship between the two variables needs to be explored in future research.

contact measure into 2.6. In the third and fourth set of models, I add the market level HHI, and then its interaction with the multimarket contact measure, respectively. In Tables 2.8 and 2.9, I start with the model with the multimarket contact measure and then add market level HHI, and the interaction term.

The results in Table 2.7 indicate a downward sloping demand curve where the coefficient for P, a_1 is negative and statistically significant. Similarly, the price of substitute for bank loans, proxied by the yield rate on the 10-year government bond is positive and statistically significant. The estimated own-price elasticity at the sample mean[32] is -1.87. That is, the estimated own price demand elasticity is in the elastic portion of the demand curve consistent with profit maximizing behavior. In the demand equation, the estimated coefficients for state net domestic product, a_3 is not significant statistically nor in economic terms as they are close to zero. The estimated coefficients for total number of branches is statistically significant but insignificant in economic terms as the estimated coefficients are close to zero. In case of the equation representing the equilibrium condition, the estimated coefficients for the cost of inputs is positive and significant as one would expect. The coefficient for output, b_{QQ} is positive but not statistically significant in any of the specifications. The positive coefficients for the time trend variable in the demand equation (a_6) indicate that the size of the loan market has increased during the sample period.

The estimated coefficients in Tables 2.8 and 2.9 are similar to the ones in Table 2.7. We find demand is inelastic irrespective of the multimarket contact measure used. The coefficient for substitute to bank loans remains positive and statistically significant across specifications of alternative measures[33]. The coefficients in the equilibrium condition functions follow similar pattern as in Tables 2.7.

[32]The mean for P in the sample is 10.22%.

[33]The coefficient for population and branches are negative. This could be due to the definition of the variable itself. I will use population density and branches per-capita as alternative variables in the next iteration.

Table 2.8: Results of market-level structural model - MMC2

Variable		MMC_2- Model 1		MMC_2- Model 2		MMC_2- Model 3	
		Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
P	a_1	-.183***	(.048)	-.125***	(.047)	-1.098***	(.103)
10-yr govt bond yield	a_2	.914***	(.076)	.789***	(.076)	2.071***	(.166)
NSDP	a_3	0	(0)	0	(0)	0	(0)
State Population	a_4	-.01***	(.002)	-.007***	(.002)	-.015***	(.005)
Branches	a_5	0***	(0)	0**	(0)	0***	(0)
T(time)	a_6	.23***	(.017)	.276***	(.017)	.245***	(.037)
Marginal Cost Constant	b_Q	1.09	(116.572)	1.143	(107.507)	1.116	(118.211)
ln(Q)	b_{QQ}	.063	(15.595)	.072	(14.382)	.053	(15.815)
$\ln(\frac{w_1}{w_2})$	b_{Q1}	.538***	(.078)	.628***	(.071)	.39***	(.096)
ln T	b_{QT}	-.069***	(.02)	-.005	(.018)	-.196***	(.068)
Lambda Constant	λ_0	1.976***	(.535)	1.494***	(.571)	12.268***	(1.49)
MMC	λ_1	-.018	(.026)	-.051**	(.025)	-.516**	(.21)
HHI	λ_2			.513**	(.249)	40.131***	(11.53)
MMC*HHI	λ_3					-4.47**	(1.748)
Observations		465	465	465	465	465	465

Table 2.9: Results of market-level structural model - Using MMC3

Variable		MMC_3 - Model 1		MMC_3- Model 2		MMC_3- Model 3	
		Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
P	a_1	-.782***	(.167)	-1.989***	(.159)	-1.989***	(.161)
10-yr govt bond yield	a_2	1.301***	(.267)	3.25***	(.254)	3.251***	(.258)
NSDP	a_3	0**	(0)	0***	(0)	0***	(0)
State Population	a_4	-.013	(.009)	.008	(.008)	.008	(.008)
Branches	a_5	-.001***	(0)	-.001***	(0)	-.001***	(0)
T(time)	a_6	.801***	(.058)	.444***	(.057)	.444***	(.058)
Marginal Cost Constant	b_Q	168.8*	(86.914)	1.781	(86.394)	1.623	(106.273)
ln(Q)	b_{QQ}	-22.442*	(11.627)	-.108	(11.557)	-.086	(14.211)
$\ln(\frac{w_1}{w_2})$	b_{Q1}	.393***	(.064)	.409***	(.059)	.409***	(.065)
ln T	b_{QT}	.048***	(.017)	.043**	(.017)	.042	(.039)
Lambda Constant	λ_0	1.104*	(.604)	2.45	(1.493)	2.37	(2.666)
MMC	λ_1	66.237***	(14.67)	177.521***	(19.418)	178.231***	(30.231)
HHI	λ_2			-11.775***	(3.806)	-11.028	(25.54)
MMC*HHI	λ_3					-6.732	(217.678)
Observations		465	465	465	465	465	465

The estimated parameter is $\widehat{\lambda} = 3.773$ and it is statistically significantly different from

zero even at the 1% level of confidence. Therefore, we can reject the hypothesis of perfect competition in this market, consistent with our findings using the Lerner index and the Boone indicator. However, we cannot reject the hypothesis of perfect collusion given the estimated parameter. It is significantly greater than 1, indicating that during the sample period, the banks' perceived marginal revenue is strongly similar to the marginal revenue taken into consideration by a monopoly or a cartel.

The estimated coefficients for multimarket contact measure, MMC1, is negative across all three specifications (and significant in models 2 and 4) in Table 2.7 . The coefficients for the second multimarket measure is also negative in Table 2.8. However, the mean values for MMC1 and MMC2 in the sample are 3.35 and 5.86 respectively. Given the estimated coefficients, we still find the structural parameter for the market to be positive and significantly greater than 1. For instance, in Model 2, we have, $\hat{\lambda} = \hat{\lambda_0} + \hat{\lambda_1} * m\bar{m}c1$ which is 3.68. The negative coefficients for the MMC measure can be interpreted as increase in competitive behavior as multimarket contact increases which in turn increases the difference between banks' perceived marginal revenue from that the marginal revenue considered by a monopoly or cartel. This could be because the two MMC measures, MMC1 and MMC2, might only be capturing part of the role played by multimarket contact. Both measures are defined as pairwise measures. However, MMC3 additionally accounts for the concentration in the market. More concentrated markets are more likely to facilitate tacit collusion between banks. The coefficients for MMC3 in Table 2.9 are positive and statistically significant across all specifications[34]. This measure is more representative of the mutual forbearance theory (or the linked oligopoly theory) as it measures the average HHI of non-home markets per contact between two banks in the same market. Therefore, greater the HHI in the other non-home markets (i.e. markets other than the one under consideration), greater is the threat of retaliation. To address potential endogeneity in the multimarket contact measure,

[34]The coefficients are large in absolute magnitude because the measure itself is defined to be between 0 and 1 unlike MMC1 and MMC2 measures.

I have estimated the same specifications where this measure is instrumented it using its lagged value. The estimated results are similar to the ones reported here.

Contrary to existing results from other studies, I find evidence of cartel-like or collusive behavior between banks in the Indian banking industry. A potential explanation for this behavior is the existence of a few large banks in the industry who capture a significant proportion of the market throughout the sample period. Earlier studies that characterize the Indian banking industry as monopolistically competitive markets. However, based on the estimated results in this paper, I argue that there is a dominant bank in the industry both in terms of its size (measured as share of total assets, credit, and deposits of the entire industry) along with its large branch network[35]. The dominant bank sets the profit-maximizing price for the market and other banks in the industry determine their prices off of the residual demand curve. Furthermore, multimarket contact between banks decreases the ability of banks to competitively price their loans vis a vis the larger banks in the industry. The threat of retaliation from larger banks in other markets could be inducing smaller banks from not competing aggressively with their larger counterparts. The smaller banks might also be competing along dimensions other than price such as in terms customer service or in terms of alternative products[36].

Why did the other measures of competition such as the Boone indicator and the Lerner indices not find evidence of collusive behavior in Dharmarajan (2023)? These indicators are based on partial equilibrium analysis which involve estimating the marginal costs. The Bresnahan-Lau model is a general equilibrium model which requires the simultaneous estimation of both demand as well as the profit maximizing equilibrium conditions [37]. In case of the Lerner index, the price is an imputed price which ignores factors that can simultaneously

[35]The market power of the said bank is further enhanced by its status of being a public sector bank with a pan-India branch network.

[36]We can test this hypothesis with individual loan data by bank. Unfortunately, I do not have access to such data.

[37]In a rough sense, we are simultaneously estimating the demand curve and the supply curve in the market level model.

affect demand for loans as well the firm's ability to price their output. One such factor is the presence of non-performing loans that can increase the risk faced by banks in the output market. This not only raises the bank's costs, it also simultaneously leads to a lower imputed price as interest income on loans fall as non-performing loans increase. Boone indicator suffers from a similar issue as it is also measuring marginal costs and its relationship with market shares of individual banks in a partial equilibrium framework. Moreover, any cost function estimates can suffer from bias and consistency if there is input misallocation (Kumbhakar and Wang (2006). These measures are most likely to be plagued by allocative errors in inputs and therefore, estimates of the marginal cost used in construction of these measures.

The results reported here is in line with the findings of Das and Kumbhakar (2016), They use an input distance function approach which do not suffer from inconsistency in cost estimates even when input allocations are not optimal. The input distance function approach is dual to the cost function but does not impose directly impose cost minimizing behavior in estimating the IDF. This is particularly relevant in the context of the Indian banking industry as the public sector banks are more likely to disburse credit or open new branches at the behest of the government rather than make these decisions with an objective to maximize profits or minimize costs (or both). Das and Kumbhakar (2016) study data from 1991-2010 and find that average markup of banks decreased during the first decade after the reforms, i.e. from 1992-2002. However, since 2002, banks started losing market share but they were less likely to adjust their relative prices and thus continued to enjoy higher markups[38]. They find that larger banks, measured in terms of their total assets, enjoyed greater market power either due to cost advantages or due to their capacity to impose higher prices.

Molnar et al. (2013) also find similar results in the Italian banking industry. They study

[38]I cannot replicate the results for the same time period as Das and Kumbhakar (2016) as bank branch data is not available for years prior to 2004. There is branch level data published by the RBI but the dataset is too noisy and can create substantial measurement errors. For instance, the branch level data reflects the most recent details about the branch which makes it impossible to track branch locations across time.

the market for deposits in the retail banking industry and find that the market is best characterized by partial collusion based on multimarket contact between the participants in the market. Moreover, they find the best fitting collusive model includes 8 banks with at least 19 markets in which they meet or operate in the presence of the other. They find that banks with extensive multimarket contact are less competitive and mimic joint profit-maximization like a cartel while taking into account the competitive fringe of smaller banks.

2.6 Conclusion

In this paper, I show evidence supporting the theory of mutual forbearance (or linked oligopoly theory). Using a fixed effects model on an unbalanced panel of 110 banks in India between 2005-2021, I find that multimarket contact among these banks is positively associated with their earnings (referred to as RoA in the paper). This relationship is robust to alternative definitions of multimarket contact measure and the set of control variables used. I find that a 1 unit increase in multimarket contact is associated with an increase in a bank's RoA (return on assets defined as the ratio of earnings to assets) by 18.4 to 25 basis points, on average. I also use a structural model in a general equilibrium framework to validate the results from the fixed effects model. I show that there is evidence of collusive or monopoly-like behavior in the market for loans in the Indian banking industry.

Existing studies measuring competition in the Indian banking industry use partial equilibrium models, irrespective of whether the methods used are parametric or non-parametric. As a result these studies (Prasad and Ghosh 2006, Rakshit and Bardhan (2019)) fail to find evidence of collusive or monopoly-like behavior. However, my findings add to the evidence found by Das and Kumbhakar (2016), who find that markups on loans have been increasing over the period 2002-2012. I add to their evidence in two ways: (a) I argue that the markups have remained high because of tacit collusion facilitated through multimarket contact be-

tween banks. I find empirical evidence in support of this argument in this paper; (b) I also show that higher markups have persisted beyond 2012 and throughout the sample period in this study.

There are a few more relevant research questions to be investigated in future research. First, are there differences in how multimarket contact facilitates tacit collusion during times of economic expansion versus times of contraction? We can take advantage of the years leading upto the global financial crisis and the period between 2012-2016[39] as expansionary periods while the remaining years can be considered contractionary. Second, does the impact of multimarket contact on a banks' earnings vary by bankgroup? Ex-ante, I would expect multimarket contact to have the strongest impact on public sector banks relative to the other two types, particularly because the public sector banks are required by law to not be allowed to fail by the Government of India.

Burgess and Pande (2005a) and Young (2021) find positive effects of policies geared towards increasing branches in under-banked areas. However, my findings indicate that we have most likely unlocked the lower end of potential benefits from increasing access to banking services. The findings from this study suggest that the government policy needs to be targeted at the level of individual markets to increase access to finance and reduce the price of loans. Though it is difficult for government or the regulatory authority to break tacit collusion between banks, especially in a large market such as the Indian banking industry, it should be part of the calculus while making fiscal and monetary policy decisions. Stronger monitoring measures are required to prevent larger banks from exercising their market power through multimarket contact. This in turn should help spur further economic growth in these underbanked regions.

[39]The demonetization exercise in 2016 started off a contractionary period for the Indian economy.

Chapter 3

Revisiting the Role of Gender in Financial Inclusion: Evidence from India

3.1 Introduction

Financial inclusion is defined in terms of the use of formal financial services. There is an increasing body of research evaluating the beneficial effects for individuals, economic and political rationale calling for government policies that improves financial inclusion. Among this research, the most the most compelling evidence is regarding the use of bank accounts. For instance, Aportela (1999) finds that having a bank account on average increases savings, Ashraf et al. (2010) finds that having a bank account increases female empowerment, while Dupas and Robinson (2013) find evidence that having a bank account increases the consumption and productive investment of entrepreneurs.

As of 2021, only 76% of adults had an account at a bank or regulation institutions (Demirguc-Kunt et al., 2022). Demirguc-Kunt et al. (2022) find that despite improvements in access to bank accounts over time[1], women are 6 percentage points less likely than men to have their own bank account in low and middle income countries. Moreover, one should read into these improvements with caution as having an account is not the same as using it actively. The share of account owners with inactive accounts is particularly high in India. Around 35% of individuals in India have inactive accounts and this is seven times larger

[1]Between 2011 and 2021, account ownership increased from 51% to 76% of all adults worldwide Demirguc-Kunt et al. (2022).

than the average for all other developing economies, excluding India. This is largely because of the Jan Dhan Yojana programme launched by the Government of India in 2014. The mission statement for this programme reads: "Pradhan Mantri Jan-Dhan Yojana (PMJDY) is National Mission for Financial Inclusion to ensure access to financial services, namely, a basic savings & deposit accounts, remittance, credit, insurance, pension in an affordable manner." Under this program, more than 490 million additional Indians were brought into the formal financial system as of May 2023[2], including 272 million women. Even though the share of inactive accounts in India have remained unchanged between 2017 and 2021, (Demirguc-Kunt et al., 2022) find that 12% more Indian women who own bank accounts have inactive bank accounts relative to Indian men who own bank accounts and are inactive users. This is in stark contrast to rest of the developing economies where the rate of bank account inactivity is roughly the same between men and women. This calls for further research into understanding the usage pattern of financial services and reasons behind higher inactivity rate among women in India.

Such differences in access to financial services can have long-term impact on individuals, households and the macroeconomy. Fletschner (2008) uses a data from survey of 210 households and finds that there is a 11% loss in efficiency in terms of what the household could have produced when the women in the household reports not being able to meet their need for credit. Klasen and Lamanna (2009) find that gender equality (or its lack thereof) potentially accounts for significant proportion of the gap in economic growth rates between East Asia and the Pacific. There is increasing evidence on the benefits of having access to formal financial services (Moore et al. (2019), Jack and Suri (2014), Field et al. (2021), Prina (2015), Lee et al. (2021)). Having an account and thus, access to formal financial resources, whether with a bank or regulated institutions[3], allows the account holders to securely participate in everyday transactions, plan for emergencies, and make investments. In the absence of

[2]Details of the programme can be accessed here: `https://pmjdy.gov.in/account`.

[3]Regulated institutions include credit unions, microfinance institutions, and mobile money service providers.

an account, individuals are forced to rely on informal mechanisms, such as holding cash or borrowing from moneylenders at a much higher interest rates (relative to formal sources). These benefits manifest in different ways, from increase in household's ability to withstand financial shocks due to having access to financial services (Moore et al., 2019), to higher remittances, consumption and investment from having access to digital financial services (Jack and Suri, 2014), to driving changes in gendered norms (Field et al., 2021).

This study adds to the literature by using a large, nationally representative sample of household survey data to investigate patterns of discrimination in access and use of financial resources. It is the first nationally representative survey of financial assets of individuals and households since the launch of Jan Dhan Yojana programme in 2014. While the programme reduces the barriers to opening a savings bank account, details on use of such accounts are not available at a disaggregated level. This study is uniquely placed in terms of combining data on various types of financial resources (formal and informal finance as well as in-kind loans) and investigating the different factors driving use (or inactivity) of these accounts. The study is also unique in terms of using data from 2019, prior to COVID-19. Therefore, the results discussed in this paper help set the benchmark for evaluating changes post-COVID.

In this paper, I focus on answering three questions: First, is there evidence of any gender disparity at the individual level in having a financial account with any formal institution? Second, do gender disparities exist in access to formal and informal (cash) loans at the level of the household? Third, conditional on access to finance, are there differences in the intensity in use of formal and informal sources of debt by gender of the household head? Using data from the 77th round of All India Debt and Investment Survey (AIDIS), I find that women are 29% less likely to have a bank account than men. Indian women are also 64% less likely than men in having a credit or debit card. There is consistent evidence that households with female heads are more likely to use non-bank resources to meet their savings and credit requirements. I find that households headed by females are 6% more likely to hold deposits in

non-bank formal institutions relative to their male counterparts. Similarly, households with female heads are 15% less likely to avail credit from banks while 6% more likely to borrow from non-bank formal sources (relative to households headed by male). These findings hold in terms of intensity with which a female headed household uses various credit resources relative to households with male heads. That is, households with female heads borrow more intensively from informal sources and have more in-kind loans.

In the next section, I provide an overview of the literature on this topic, followed by description of the data used and its summary statistics. I then discuss the framework of the empirical model used, and present and discuss the results from the analysis.

3.2 Background and Summary of the literature

There are several ways in which access to financial services can affect an individual's and household's growth as well as the macroeconomy. Girma and Shortland (2008) use a theoretical model to show that when households and firms have access to financial services at affordable terms and conditions it augments a country's economic growth and development. Doepke and Tertilt (2019) finds that transfers to women are more beneficial to growth when human capital is the most important factor of production. Using a non-cooperative model of the household, they show that women will spend more on the children and invest more in human capital when transfers are targeted towards women. Using data from 1960-2000 for a large panel of countries, Klasen and Lamanna (2009) find that gender gap in education and employment considerably reduces economic growth.

Swain (2002) applies a structural model of demand and supply to investigate the credit rationing by the formal credit markets in Puri, India. They use data from a sample survey of 1052 households over 66 villages in Puri. They find that household characteristics such

as the amount of land owned by the household, land quality, household size, and other macro variables at the village level are relevant factors in determining the probability of access to formal credit. Campero and Kaiser (2013) uses data from a nation-wide survey of households, in Mexico to model the decision process. The survey includes questions on awareness about different financial resources. They leverage this information in the survey to first estimate the model of credit use with consideration set formation. Then using the household's estimated choice set, they estimate the decision process using a multinomial Logit model. They separate out the effect of household and local characteristics on financial awareness from the use of different type of credit [4]. They find an inverse relationship between household income and use of informal sources of credit and a positive relationship between income and use of formal credit. They also find that awareness and use of both formal and informal sources are differently affected by schooling, where higher levels of education is positively associated with awareness of formal resources and negatively with use of informal credit sources. Moreover, women are found to be more likely to borrow from pawnshops.

However, there are limitations in implementing structural models as well as they require detailed data at the individual and household level on awareness of various sources of credit and the intensity of use. Data on outcomes where a household applied for credit but were denied is also essential. These are necessary to construct the choice set faced by the household as well as the outcomes in the financial market for each household. Even though there are multiple national survey datasets in India [5], none of the publicly disseminated data of national surveys collect data on both these aspects[6]. In this paper, we use data from AIDIS and we are unable to apply a structural model of demand and supply of formal loans. We are able to only observe outcomes where a household succeeded in obtaining a loan from a

[4]AIDIS survey data, which I use in this study, unfortunately does not collect data on awareness and limits the scope of the analysis in the current paper.

[5]This includes AIDIS, yearly rounds of National Sample Surveys covering different theme each year, National Family Health Survey, District Level Household Survey, Indian Human Development Survey.

[6]The Indian Human Development Survey collects does collect information on various loans that a household had applied for and its outcome. However, the survey covers substantially smaller number of households compared to AIDIS.

formal financial institution. We are unable to observe the cases where the household applied for loans from such institutions but were denied. Hence, in this paper, we use reduced form empirical models to determine the significance of various individual, household and macroeconomic factors that determine access to finance (formal and informal) and its usage.

The findings from the structural models are supported by evidence from empirical studies. There is growing literature on the impact of specific interventions fostering use of financial instruments to save at the household level. Most of these studies use randomized control trials (RCTs) given the limitations of observational data to estimate causal impact. Ashraf et al. (2006) looks at the impact of commitment devices on household savings in . Karlan et al. (2016) studies the impact of sending reminders and finds that sending these reminders increased commitment and meeting savings goals, especially for customers who had opened accounts recently. They argue that reminders help mitigate under-saving due to limited attention to exceptional expenses. Schaner (2018) evaluates the impact of temporary saving subsidies among individuals in Kenya and find that participants who received large temporary interest rates on their bank account had significantly more income and assets in the long-term even after the expiry of such interest subsidies. Kast et al. (2012) looks at how effective peer pressure is among micro-entrepreneurs in Chile in promoting use of bank accounts. They find that peer pressure has a significant impact as number deposits increased by more than 3.5 times while average balance in savings account increased by more than twice. Banerjee and Duflo (2014) use loan data availed by 253 small and medium sized firms from an Indian bank and find that after the firms obtained access to credit, they were able to expand their businesses. McKenzie and Woodruff (2008) found that receiving grants to procure inputs led to higher returns for micro-enterprises in Mexico using an randomized control trial. Klapper et al. (2006) finds that access to financial services is a key driver of entry of new firms and such access supports growth of small firms. Though these studies identify various interventions that can promote use of bank accounts and measure its impact cleanly, they suffer from a few limitations. RCTs cannot be used as a methodology to study

impact of macroeconomic policies and institutional country-wide features. Moreover, RCTs can also suffer from external validity as certain interventions may not scale at a national level.

Badarinza et al. (2016) finds balance sheets of Indian households to be distinct from those in other economies along three key aspects. First, they find, relative to households in other economies, the average Indian household allocates higher share of their wealth to physical assets, particularly in gold and residential real estate. Second, the significant fraction of household debt is sourced from non-institutional source. Moreover, there is a high reliance on securing these debts with gold in certain states. Third, they find Indian households have lower stock of retirement assets. This study is based on the 70th round of AIDIS survey. They find that place of residence of household, age and education of household head, number of children, and wealth are relevant factors in determining households' portfolio of debt and investment. Similarly, Rampal and Biswas (2022) finds that higher educational attainment, degree of social connectivity and confidence in financial institutions to be key predictors of households portfolio of assets. However, they use data from another national survey - Indian Human Development Survey (IHDS-2, 2011-12). Govindapuram et al. (2023) uses data from the fourth round of the National Family Health Survey (2015-16) which collected information on women's ownership and use of bank or savings account. The survey also asked respondents whether they were aware of any local micro credit programmes and whether they ever took a loan from such a program for their business[7]. They find that women in male-headed households are less likely to own a bank account. Kumar et al. (2019) find that availability of banking services have negligible effect on households' use of formal services relative to other factors such as education, income, employment , gender and social norms. The latter set of factors affect the demand for formal services and they argue that we should place greater emphasis on resolving demand-side barriers to promote financial inclusion in

[7]However, this study is limited because the measure of awareness and use of loans only pertain to microfinance programs, and in particular, business loans. Therefore, the estimated effects in are with respect to business loans only and cannot be generalized across different types of loans.

India. They also find evidence of discrimination based on gender and caste in the rural areas. Households headed by female head and households belonging to lower castes are more likely to be financial excluded. Pal (2022) use a survey of 600 households from 14 villages spread over 6 states and find that households that avail multiple loans from the formal sector are those that reside in villages close to critical infrastructure such as educational institutions and agricultural infrastructure. Pal (2022) also finds that households belonging to upper social strata, and being a Hindu household, having assets that can be mortgaged, operational land, and have regular employment are more likely to avail multiple loans from formal sources.

In this paper, I use most recent data on household assets, investments and debt, to examine the relevance of numerous household level characteristics in determining an individual's and household's access to financial resources, and their use. This work adds to the literature in two ways. First, the key contribution of this study is in terms of combining both savings and borrowings of an household across various sources. Past studies have focused on either savings or on borrowings (credit) due to limitations in the data. I utilize the 77th round of the AIDIS survey data to empirically examine differences in access to finance by gender, both at the individual level and household level. This is also one of the first studies to use a large, nationally representative sample of household survey data to investigate patterns of discrimination in access and use of various types of financial resources (formal and informal finance as well as in-kind loans). Second, the data used in this study is timely. The survey data was collected in 2019, prior to the beginning of the COVID-19 crisis. It thus helps evaluate the benchmark before a structural break in the economy. At the same time, as the survey covers data from 2018-19, we can use this to evaluate the progress of the Jan Dhan Yojana programme.

3.3 Data Sources and Summary Statistics

The primary data set used in the analysis is the 77th round of All India Debt and Investment Survey (AIDIS) conducted by National Sample Survey Organization (NSSO). NSSO is part of the Ministry of Statistics and Implementation (MOSPI). This a nationally representative survey that collected information at both individual as well as at the level of the household on a wide variety of features. The survey uses a two-stage stratified sampling design for both rural and urban areas. While villages or sub-units of villages are the first stage units, blocks or sub-units of blocks are the first stage units in urban areas [8]. After selecting the first stage units using simple random sampling without replacement, households are selected in the second stage within each first stage unit. The households are also selected using Simple Random Sampling without replacement. There were a total of 69,455 rural households and 47,006 urban households surveyed [9]. The sample of households in this analysis consists of 111, 508 households. The sample of individuals on the other hand consists of 495,573 individuals. However, only 289,512 individuals are older than 21 years of age. The survey consists of two rounds The survey collected data on all assets (both financial and non-financial) and liabilities (cash and in-kind) of households as of 30-June-2018. The first visit of the survey was completed between January-August 2019 while the second visit was completed between September-December 2019 [10]. As the 77th round of the AIDIS survey data completed before March-2020, we can avoid the confounding effects of COVID-19 on household's access to and use of financial resources.

The 77th round of the AIDIS data collects data on ownership of land at the individual level and at the household level. The questionnaire is broken down into 14 blocks and it collects

[8] A total of 5,940 rural villages and 3,995 urban villages were selected as first stage units.

[9] However, we dropped 2,026 rural households and 2,827 urban households as the respondent was marked as either busy or reluctant at the time of the survey.

[10] We focus our analysis on the first round of the survey as there are fewer households surveyed in the second round because some households could not be survey for one reason or another.

information on demographic characteristics of the household as well as its members, the different assets owned by the household such as land, buildings and constructions, livestock and poultry, transport equipment, agricultural implements, non-farm equipment, financial, non-financial and investment assets, loans (both in cash/credit as well as in-kind) , as well as transactions conducted by the household (expenditures on purchase, additions, renovations etc.). The first visit of the survey collects information on all 14 blocks while the second visit of the survey collects information only for Blocks 1, 2, 12, 13 and 14. While individual and household characteristics are defined using data from AIDIS, various state level features are defined using data from the Handbook of Statistics on Indian States published by the Reserve Bank of India[11].

The main features are defined at the individual and at the household level. At the individual level, there are three key features - whether individual has any deposit account, whether individual holds any debit or credit card, and whether individual owns any land. All three are binary variables where we classify an individual to have a deposit account if they have a deposit account either in commercial banks or regional rural banks, cooperative banks, post office or non banking financial companies. Having a deposit account in these sources indicates that the household has access to savings with formal sources of finance. For the individual level analysis, we use whether individual has any deposit account and whether individual holds debit or credit cards as outcome variables while including whether individual owns any land as one of the control variables. The summary statistics for these features are presented in Table 3.1 Panel B. However, we do not have data on neither deposit nor loan/credit at the individual level. At the individual level, we can only examine access to financial savings instruments but not its usage. We cannot examine access and use of credit either at the individual level either.

We use features defined at the household level to evaluate use of various financial services.

[11]Table A.1 in Appendix A has details of each feature and its source.

Households are classified into those where the head of the household identifies as male and those where they identified as female. This allows us to study whether there is gender discrepancies in use of financial resources. A household is defined to have access to formal finance if they have availed atleast one cash loan from a formal source, where formal source includes banks and non-bank sources. Formal sources of credit includes credit from scheduled commercial banks, regional rural banks, cooperative societies, cooperative banks, insurance companies, provident fund, financial corporation/institutions, Non-Banking Financial Companies (NBFC) and Micro-Finance Institutions (MFI), and bank and non-bank linked Self-Help Groups (SHGs) and Joint Liability Groups (JLGs). Informal sources of credit on the other hand includes landlord, agricultural moneylenders, professional moneylenders, input suppliers, relatives and friends, chit funds, and market commission agent/traders. For every cash loan taken by the household, we use the original borrowed amount[12]. A household is defined to have access to instruments to save from formal sources related to a bank if they have a current account at a bank or have deposits in a savings bank account or deposits in fixed deposit, term deposit, recurring deposits, or flexi-recurring deposits in banks. A household has access to instruments to save from formal and informal resources that are non-bank if they have a deposit in other fixed income deposits (such as National Savings Certificate, Kisan Vikas Patra, savings bonds, other small savings schemes), deposits in cooperative banks, NBFCs, co-op credit societies, MFIs, and SHGs.

Table 3.1 reports the mean and standard deviation of these features aggregated over all individuals and households, and by gender. At the individual level, 83% of all individuals in the sample (N=289,512) who are older than 21 years old report that they have a bank account. However, once we break it down by gender, we see that while 89% of men (N=146,380) have a deposit account, only 77% of women (N=143,132) have a deposit account. Only 42% of individuals in the sample report having a debit or credit card but there is a clear disparity

[12]The survey collects information on original amount borrowed, amount outstanding as of 30-June-2018 and the amount outstanding as of the date of survey.

by gender for having debit or credit cards. While 54% of men in the sample have a debit or credit card, only 30% of women report to have one. There also exists stark differences in land ownership among individuals. Across the sample of individuals, 33% report owning any land with 55% of men in the sample owning land as compared to only 11% of women owning any land. Evidence of such disparities in access to and use of financial resources also exist at the household level.

For an overview of the household's access to and use of credit from various sources, we look at the median value of the feature for the median household.[13] Within the sample of households who have availed at least one cash loan [14], the number of different cash loans availed by the household is 1. For both types of households (those headed by a male and those by a female) have one loan (at the median) from all sources (formal and informal). Breaking this down by formal and informal sources, the number of cash loans from formal sources for the household is also 1 loan per household. The median number of loans from formal sources is one for both households headed by a male and by a female. The same is true for loans from informal sources as well. In terms of intensity of use of credit, the median loan size among all credit availed by a household from both formal and informal sources is 70,000 INR (= 875 USD). The same statistic is 50,000 INR (= 625 USD) per household headed by a woman and 70,000 INR (= 875 USD) per household headed by men.

[13] We look at the median instead of the average to minimize the bias from extreme values.

[14] There are 69,138 households in the sample who have availed at least one loan from either a formal or an informal source. Of these households, only 11, 425 households have borrowed from both formal and informal sources. 34,795 households have only borrowed from formal sources while 22,918 households have only borrowed from informal sources.

Table 3.1: Summary statistics of main outcome variables

PANEL A: Individual level features						
Variable Name	All (N=289,512)		Males (≥21 years) (N=146,380)		Females (≥ 21 years) (N=143,132)	
	Mean	StdDev	Mean	StdDev	Mean	StdDev
Individual's age	42.09	15.04	42.22	15.09	41.95	14.98
Individual has deposit account (binary)	0.83	0.37	0.89	0.31	0.77	0.42
Individual holds debit or credit card (binary)	0.42	0.49	0.54	0.50	0.30	0.46
Individual Owns Land (binary)	0.33	0.47	0.55	0.50	0.11	0.31
PANEL B: Household level features related to access						
Variable Name	All households (N=111,508)		Households head by males (N=96,567)		Households headed by females (N=14,941)	
	Mean	StdDev	Mean	StdDev	Mean	StdDev
Household has deposit account (binary)	0.86	0.34	0.91	0.28	0.54	0.50
Household has debit or credit card (binary)	0.57	0.49	0.60	0.49	0.38	0.49
Household owns land (binary)	0.78	0.42	0.83	0.38	0.47	0.50
Number of member in household with deposit account	2.16	1.42	2.28	1.36	1.40	1.55
Number of members of household with debit and credit card	1.09	1.24	1.14	1.24	0.75	1.17
Number of household members owning land	0.85	0.56	0.91	0.53	0.53	0.63
Total amount of cash held by household (INR)	4106.85	13033.78	4254.75	13275.17	3150.93	11304.27
Total amount of deposits held in Banks by household (INR)	41649.98	347941.44	43234.99	367549.98	31405.71	173976.13
Total amount of deposits held in Non-Bank formal sources by household (INR)	2524.08	32987.49	2563.42	33693.09	2269.76	28001.09
Household has atleast one loan from formal sources (binary)	0.41	0.49	0.43	0.49	0.32	0.47
Household has atleast one loan from bank (binary)	0.27	0.44	0.28	0.45	0.18	0.39
Household has atleast one loan from non-bank formal sources (binary)	0.14	0.35	0.14	0.35	0.13	0.34
Household has atleast one loan from informal sources (binary)	0.31	0.46	0.31	0.46	0.28	0.45
Total amount of formal loans availed by household (INR)	123854.95	595284.61	131317.88	616754.75	75620.44	428370.18
Total amount of informal loans availed by household (INR)	30922.49	128277.16	32129.43	133086.68	23121.80	90889.54
Household has atleast one loan in-kind (non-cash) (binary)	0.07	0.25	0.07	0.25	0.06	0.24
Total value of in-kind loan availed by household (INR)	905.86	12436.01	949.16	12854.38	625.98	9283.58

On breaking down the size of credit by formal and informal sources, the median loan size from a formal source is 80,000 INR (= 1000 USD), while the median loan size from an informal source is 0 [15]. Furthermore, a t-test of means of the amount borrowed (from both formal and informal sources) between households headed by a male and that by a female provide evidence of statistically significant differences between the two. The average amount of credit for an household headed by a male is 256,978 INR (= 3212 USD), while the average amount of credit with households headed by a female is 191,152 INR (= 2,389 USD). The average amount of credit from formal sources for a household headed by a male is 206,643 INR (= 2583 USD), while the average amount of credit with households headed by a female is 146,391 INR (= 1830 USD). A t-test of the means indicate that households headed by a male have availed significantly more in credit from formal sources relative to households headed by female (t-stat=8.16). Similarly, the average amount of credit from informal sources for an household headed by a male is 50,515 INR (= 631 USD), while the average amount of credit with households headed by a female is 44,761 INR (= 560 USD). On running a t-test of the means, we find households headed by a male also avail significantly more in credit from informal sources relative to households headed by female (t-stat=3.73). Household's seem to compliment these sources of credit with in-kind loans.

From a qualitative study of 200 households, Basavaraj and Bhattacharjee (2014) finds that loans from informal sources are relatively smaller and are primarily used for current expenditures. While loans from formal sources are not only larger and lumpier, they are more likely to be used for capital expenditures. We also observe that in terms of the amount borrowed, the median amount borrowed from formal sources is greater than the median amount borrowed from informal sources for all types of households. Within the sample of households in our analysis, we find that 65% of loans from informal sources are used to meet

[15]The median loan size from a formal source for households headed by a male is 88,000 INR (= 1,100 USD) while the same for households headed by a female is 60,000 INR (= 750 USD). The same set of numbers for credit from informal sources are 40,000 INR (= 500 USD) across all households, 40,000 INR (= 500 USD) for households headed by a male, and 30,000 INR (= 375 USD) for households headed by a female.

other household current expenditures (including medical expenditure) whereas 54% of loans from formal sources are used to meet capital expenditure in business and for investments such as in housing.

Table 3.2: Summary statistics of other household characteristics

	All households (N=111,508)		Households head by males (N=96,567)		Households headed by females (N=14,941)	
	Mean	Std	Mean	Std	Mean	Std
Age of household head	48.45	13.92	47.87	13.67	52.15	14.87
Size of the household	4.26	2.16	4.41	2.12	3.29	2.15
Number of children in household	1.39	1.40	1.45	1.41	1.03	1.29
Dummy = 1 if highest education level in household is illiterate or below primary	0.11	0.31	0.09	0.29	0.21	0.41
Dummy = 1 if highest education level in household is primary and above, upto higher secondary	0.63	0.48	0.64	0.48	0.57	0.49
Dummy = 1 if highest education level in household is graduate and above	0.26	0.44	0.27	0.44	0.21	0.41
Dummy = 1 if household's religion is Islam	0.12	0.32	0.12	0.32	0.11	0.32
Dummy = 1 if household's religion is Christian	0.06	0.24	0.06	0.24	0.08	0.27
Dummy = 1 if household's religion is Others (not Hindu)	0.04	0.19	0.04	0.19	0.04	0.19
Dummy =1 if household's caste is Scheduled Caste	0.16	0.37	0.16	0.37	0.18	0.38
Dummy =1 if household's caste is Scheduled Tribe	0.15	0.35	0.15	0.35	0.15	0.36
Dummy =1 if household's caste is Other Backward Caste	0.41	0.49	0.41	0.49	0.40	0.49
Dummy =1 if household's caste is Others	0.29	0.45	0.29	0.45	0.28	0.45
Dummy =1 if household's employment is Regular Salaried	0.22	0.41	0.22	0.41	0.19	0.40
Dummy = 1 if household's employment is Casual Labor	0.23	0.42	0.23	0.42	0.22	0.41
Dummy = 1 if household's employment is Self-employed	0.47	0.50	0.49	0.50	0.32	0.47
Dummy = 1 if household's employment is Others	0.09	0.28	0.06	0.24	0.27	0.44
Dummy = 1 if household's residence is Urban	0.40	0.49	0.39	0.49	0.42	0.49
Household's usual monthly total expenditure (INR)	11185.56	8298.28	11510.43	8419.70	9085.85	7117.16
Value of households immovable assets (INR)	2128959.30	8001805.98	2211561.02	8012128.67	1595086.02	7914263.57
Value of land owned by households in rural area (INR)	989997.61	6160567.70	1050905.69	6505848.85	595563.97	3074378.70
Value of land owned by households in urban area (INR)	479619.60	3314906.89	491637.88	3456024.72	401884.84	2191302.67
Value of buildings owned by household (INR)	660091.22	3359444.29	669482.09	2506257.32	599395.96	6605274.53

Table 3.2 includes summary statistics of other household characteristics. Household headed by a female are smaller in size and have lower levels of education (in terms of the highest educated member). Moreover, households head by a female have statistically significantly lower expenditure per month and lower amount of immovable assets [16]. As a household's wealth position improves (measured by value of immovable assets owned), their access to credit should improve as well, given constraints on providing collateral reduces. The average value of immovable assets for households headed by a male (2,211,561 INR or 27,645 USD) is economically and statistically significantly greater than the average value for households headed by a female (1,595,086 INR or 19939 USD) [17]. We find a similar difference in terms of monthly household expenditure on usual monthly consumption expenditure, imputed value of usual consumption from home-grown stock, wages in-kind, free collections, gifts etc, and expenditure on durable goods. The average monthly expenditure for households headed by a male is 11,510 INR (or 144 USD) compared to 9086 INR (or 114 USD) for households headed by a female [18]. In our sample, households headed by women have fewer immovable assets and have lower monthly expenditure relative to households headed by men. We also include additional features at the state level that can potentially affect a household's access to finance. This includes supply-side factors for access to finance, such as the number of bank branches in the state, number of bank branches in the state managed by women, bank credit, and bank deposit in each state.

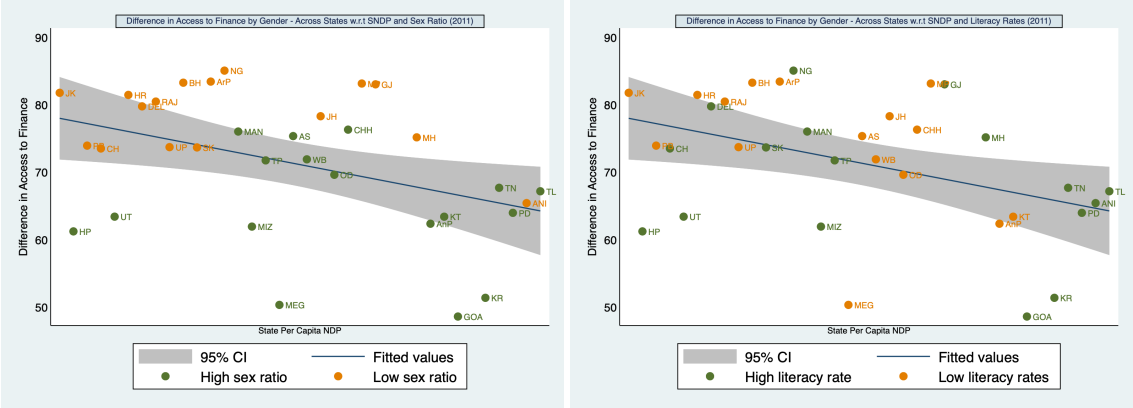
We also include other social factors that can affect demand for credit such as State Net Domestic Product, various measures that proxy for infrastructure in the state (such as length of railways and roads in the state, power per capita, population by rural and urban areas in the state, enrollment rates in primary and secondary schools as well as the rate of crime

[16] We classify households into quartiles based on their monthly expenditure and immovable assets and include the quartiles as control features. We classify households into quartiles based on the distribution of the feature within the state. This allows us to account for potential differences in economic conditions that can affect the value of these features.

[17] The t-stat from a difference in means test is 8.85.

[18] The t-stat from a difference in means test is 37.75.

Figure 3.1: Differences in access to formal finance



(a) Difference in access to formal credit - by Gender of household head, NSDP, and Sex Ratio

(b) Difference in access to formal credit - by Gender of household head, NSDP, and literacy rate

against women in the state). For instance, Panel A of Figure 3.1 plots the proportion of number of households headed by males in the given state who have atleast one cash loan from a formal source to number of households headed by females in that state who have atleast one cash loan from a formal source, on the y-axis. Greater the proportion, greater the degree of discrepancy in access to formal credit by gender of the head of the household. The x-axis represents the State Net Domestic Product (SNDP) at constant prices for 2018. The states are arranged by their SNDP. The states are further classified on the basis of whether they are above or below the median sex ratio across all states. Panel A of Figure 3.1 shows that in states with sex ratio below the median (orange) there are substantially greater number of households headed by male that have access to formal credit relative to households headed by females. Similarly, Panel B of Figure 3.1 shows that states with literacy rates below the median literacy rate (orange) have higher differences in access to formal credit, net of controlling for NSDP. Therefore, state-level macro variables could be relevant for a household's demand for financial services and supply of such services in the state.

Households can also face discrimination in accessing credit through the interest rates charged.

Table 3.3: Summary of interest rates on credit - by interest type, gender of household head, and place of residence

Nature of interest rate	All				Rural				Urban			
		Household headed by male	Household headed by female	T-test of means (t-stat)		Household headed by male	Household headed by female	T-test of means (t-stat)		Household headed by male	Household headed by female	T-test of means (t-stat)
All	Obs	31,394	3,453	-6.77***	Obs	18,439	1,843	-5.15***	Obs	12,955	1,610	-3.23***
	Mean interest rate	10.91	11.53		Mean interest rate	10.35	11.02		Mean interest rate	11.71	12.12	
Simple	Obs	12,575	1,397	-4.99***	Obs	8,580	893	-3.73***	Obs	3,995	504	-2.61**
	Mean interest rate	10.20	10.95		Mean interest rate	9.65	10.33		Mean interest rate	11.38	12.04	
Compound	Obs	18,819	2,056	-4.74***	Obs	9,859	950	-3.87***	Obs	8,960	1,106	-2.07**
	Mean interest rate	11.37	11.93		Mean interest rate	10.95	11.68		Mean interest rate	11.85	12.15	

Note: *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

Tables 3.3 and 3.4 show that households headed by women pay significantly (statistically) higher interest rates on their loans (both formal and informal) compared to households headed by men. The survey collects information on both the interest paid by the household and the type of interest charged, i.e. whether it is simple interest or compound interest. Considering the loans where information of interest type is not missing, in Table 3.3, we find that households headed by females pay higher interest rates relative to households headed by males, irrespective of whether the household resides in an urban area or rural area, for all types of interest charged (simple and compound). The t-statistics from the difference in means of interest rates paid by gender of household head is provided in the table as well. We can also see that interest rate paid by households in urban areas are higher than in rural areas.

Table 3.4 examines the differences in interest paid by household type (whether headed by male or female), type of interest paid, and by asset quartile (based on value of immovable assets held by household). We find that for households in the lowest quartile, there is no difference in interest rates paid by gender of the household head. In the lower asset quartiles, the lack of collateral for both types of households could be a factor in determining the interest rates charged. As a result, it will be harder to find evidence of gender based discrimination in interest rates charged for credit. However, in the highest quartile, we find that households headed by female pay higher interest on their loans relative to households headed by male. That is, despite having considerable immovable assets to provide as collateral for both types

Table 3.4: Summary of interest rates on credit - by interest type, gender of household head, and asset quartiles

Nature of interest rate	Quartile 1				Quartile 2			
		Household headed by male	Household headed by female	T-test of means (t-stat)		Household headed by male	Household headed by female	T-test of means (t-stat)
All	Obs	3,916	666	-0.17	Obs	7,210	932	-0.51
	Mean interest rate	12.75	12.80		Mean interest rate	11.85	11.95	
Simple	Obs	1,358	248	-0.04	Obs	3,027	411	-1.01
	Mean interest rate	12.01	12.03		Mean interest rate	11.16	11.45	
Compound	Obs	2,558	418	-0.32	Obs	4,183	521	0.07
	Mean interest rate	13.15	13.26		Mean interest rate	12.35	12.33	
	Quartile 3				Quartile 4			
		Household headed by male	Household headed by female	T-test of means (t-stat)		Household headed by male	Household headed by female	T-test of means (t-stat)
All	Obs	8,887	939	-2.73**	Obs	11,381	916	-5.08***
	Mean interest rate	10.68	11.13		Mean interest rate	9.85	10.61	
Simple	Obs	3,610	375	-2.26**	Obs	4,580	363	-3.04**
	Mean interest rate	10.03	10.68		Mean interest rate	9.17	9.91	
Compound	Obs	5,277	564	-1.57	Obs	6,801	553	-4.08***
	Mean interest rate	11.12	11.43		Mean interest rate	10.32	11.07	

Note: *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

of households, households headed by female are charged higher interest rates than households headed by male. We include the interest rate paid by the household as one of the controls while estimating the intensity of use of credit by households.

We include a wide range of state level features to control for macro-level. A household's distance to the nearest branch is a determinant of their ability to use the bank's services, with increasing distance implying increasing opportunity cost for the household to access and use these bank services. However, the survey data does not include the geo-location and we cannot define household's distance from nearest bank branch. We instead include the number of bank branches in the state as well as number of branches managed by a female manager in the state as controls. Inclusion of number of bank branches as a control variable is motivated by Burgess and Pande (2005b). They find significant reduction in poverty caused in rural unbanked locations due to bank branch expansions. They find that it was mainly through increased deposit mobilization and credit disbursement by banks in rural areas.

In the next section, we discuss the main empirical specification to check for evidence of

gender disparity in access to finance and its use, both at the level of the individual and at the household.

3.4 Empirical Model

We use multivariate Probit regression model to study the determinants to access to finance. For models defined at the level of individuals, the regression specification takes the form:

$$y^*_{ihs} = X'_{ihs}\beta + X'_{hs}\theta + X'_s\gamma + \varepsilon_{ihs} \tag{3.1}$$

$$y_{ihs} = 1 \text{ if } y^*_{ihs} > 0$$

$$y_{ihs} = 0 \text{ if } y^*_{ihs} \leq 0$$

where i indexes an individual, h indexes the household to which the individual belongs to, and s indexes the state (or district) to which the individual i and household j belong to[19]. Therefore, we include both household level characteristics, and state level controls as additional explanatory variables in the model. y^*_{ihs} is a latent variable while X_{ihs} are a vector of individual specific characteristics, X_{hs} is a vector of household specific characteristics, and X_s are state-specific (or district specific) characteristics. Here we assume ε_{ihs} is a normally distributed error term with zero mean and variance equal to 1. We estimate this model in terms of Probit model (using maximum likelihood) [20].

After estimating Model 1, we report the marginal effects in Tables 3.5 to discuss the economic significance of various characteristics in access to finance along various dimensions. These

[19]This empirical specification is consistent with past literature ((Demirguc-Kunt et al., 2013), (Demirguc-Kunt et al., 2014), (Demirguc-Kunt et al., 2015), (Allen et al., 2016) and Ghosh and Dharmarjan (2017)).

[20]We also estimate the marginal effects using Logit model with state fixed effects and district fixed effects (in place state level control variables). The results are robust to these alternative specifications. We do not include fixed effects in the Probit model as it leads to biased estimates. Therefore, we instead run Logit with fixed effects when required.

marginal effects are cannot be interpreted as causal effects but as significant (or insignificant correlations) between the characteristics included and measures of financial inclusion. These marginal effects allow us to evaluate whether significant differences exist along certain dimensions of interest to us, mainly gender of the individual (for individual level regressions) and gender of the household head (at the household level).The marginal estimates can be interpreted as the probability of a change in dependent variable when a one unit change in the value of the regressor, while holding all other variables constant.

We also use a Probit model specification when the unit of observation is a household. We have:

$$y^*_{hs} = X'_{hs}\beta + X'_s\gamma + \varepsilon_{hs} \tag{3.2}$$

$$y_{hs} = 1 \text{ if } y^*_{hs} > 0$$

$$y_{hs} = 0 \text{ if } y^*_{hs} \leq 0$$

where h indexes the household to which the individual belongs to, and s indexes the state (or district) to which the individual i and household j belong too. Therefore, we include both household level characteristics, and state level controls as additional explanatory variables in the model. y^*_{hs} is a latent variable while X_{hs} are a vector of individual specific characteristics, and X_s are state-specific (or district specific) characteristics. We continue to assume that ε_{hs} is a normally distributed error term with zero mean and variance equal to 1. Similar to the model for individuals, we estimate this model in terms of Probit model (using maximum likelihood) and a Logit model when we include fixed effects for states. We report the marginal effect of the estimates in Tables 3.6 and 3.7 and these can be interpreted similarly as explained before.

To study whether gender disparities exist in use of credit (formal, informal, and in-kind), I apply Cragg's Double-Hurdle model ((Cragg, 1971)) . The main outcome of interest is the

amount of credit accessed by a household, either as cash loans from formal and informal sources or in-kind. There are two aspects of the outcome variable that necessitates the application of the Double-Hurdle model. First, the outcome variable, amount of credit from any source, is never negative. Second, there is potential selection bias that needs to be addressed while estimating the model as only the households who have utilize credit from various sources can be different from households who never utilize credit. We have a corner solution model is one where the dependent variable, y_i is:

$$y_i = y_i^* \text{ if } y_i^* > 0$$

$$y_i = 0 \text{ if } y_i^* \leq 0$$

and

$$y_i^* = \alpha + X_i\beta + \varepsilon_i$$

i.e. it is truncated and has a significant peak at a certain given value while being continuously distributed otherwise. This is the case with amount of credit used by a household and therefore a bi-variate Probit is not the appropriate model to use. Of the 111,508 number of households in the sample, only 41.4% have any positive cash loans from formal sources of credit, i.e. 58.6% of observations are equal to 0 for the main outcome variable, amount of credit. A Tobit model specification could resolve the both issues. However, the limitation of the Tobit model is that the probability of a positive value and the actual value conditional on it being positive are determined by the same parameters. This imposes a restriction that the direction or sign of a given independent variable's marginal effect on the outcome variable will be same in both tiers, i.e. the marginal effect on the probability that $y > 0$ is the same as the conditional expectation of y. Cragg (1971) provides an alternative to Tobit allowing the outcomes to be determined by different processes. The Double Hurdle model uses a

Probit model in the first step when the outcome variable is binary (whether the outcome variable in the second step will be positive or no) and uses a truncated normal model in the second step [21].

To apply the Double Hurdle model, we define the outcome variable as a binary variable, i.e. whether household h has access to formal (or informal or in-kind) credit. This binary variable is equal to 1 if household h has atleast one loan from the given source, and zero otherwise. In the second stage, we define the outcome variable as the total amount of credit availed by the household from the source in question (formal, informal or in-kind). The second stage of the Double Hurdle model is only estimated for households who have access to given form of credit in the first stage, i.e. the amount of credit availed from the given source is greater than 0. However, the estimated coefficients cannot be interpreted directly. Instead, we need to use the artifacts of the estimation results to derive the average partial effect as described in Burke (2009).

3.5 Results and Discussion

We discuss the model estimates in this section. We will start with individual level analysis and then look at household level results.

3.5.1 Individual-level findings

At the individual, we look at two outcome variables: (1) whether individual has a bank account, and (2) whether individual has a debit/credit card. To assess the different factors

[21] However there is an additional assumption on the conditional independence of latent variable's distribution, i.e. $D(y^*|w,x) = D(y^*|x)$ where $w = 1$ if $y > 0$ and 0 otherwise. As a result, the unbiasedness of estimates are sensitive to model misspecification. I address this issue by verifying the estimated effects are robust across specifications.

determining these outcomes, we apply a Probit model (Model 1) and report the marginal effects in Table 3.5. We cluster the standard errors at the household level as households are the unit of randomization in the sample and these are reported in parentheses. In Table 3.5, columns A and B take "whether individual has a bank account", a binary variable, as the outcome variable while columns C and D takes "whether individual has a debit/credit card" as the outcome variable. Columns A and C report the marginal effects from a Probit model specification with individual, household and state-specific characteristics. Columns B and D on the other hand report the marginal effects from a Logit model specification with individual and household characteristics, and state fixed effects. Comparing results between columns A and B, and between C and D, shows that the estimated effects are robust to alternative model specifications. The results in Column A and C are qualitatively similar to the results in column B and D. We will focus the discussion on the estimates from the Probit model.

We find that a female is 26% less likely to have a bank account and 64% less likely to have a debit or credit card, both relative to a male. These effects are statistically significant as well as economically meaningful. Moreover, the probability of having an own bank account and a debit or credit card increases if the individual owns land within the household. This indicates that ownership of assets provides greater bargaining power to the individual and therefore greater agency and independence. The probability of having an account and possessing a debit/credit card is higher when other members in the household have a bank account and have a debit/credit card respectively. Also, individuals in households that identify as Muslim and Christian are less likely to own a bank account relative to individuals in households that identify as Hindu. Similarly, individuals belonging to the Scheduled Tribe are less likely to have a bank account as well as have a debit or credit card compared to individuals in the general caste. We also find that individuals in urban households are less likely to have a bank account but are more likely to be using a debit or credit card relative to individuals in rural households. This could be because most transactions in rural areas are still predominantly

transacted in cash and in-kind and therefore, use of debit and credit cards are less likely. Though not reported in Table 3.5, we also find that probability of having a bank account increases with the individual's level of education[22]. The same pattern, i.e. individuals with higher education levels are more likely to have a debit or credit card, relative to an illiterate individual.

3.5.2 Household-level findings

At the household level, we start with binary outcome features that measures access to various financial services. Table 3.6 reports the marginal effects from Probit estimation in columns A and C, and marginal effects of Logit estimates in columns C and D. The outcome features are: (1) whether household has deposit with a bank, and (2) whether household has deposit with non-bank but formal sources [23]. The main variable of interest is the marginal effect on access when the head of the household is a female (Dummy = 1 if hh head is female). Similar to the discussion of individuals, results in columns B and D are qualitatively consistent with results reported in columns A and C, thus indicating that these results are robust to alternative specifications. We will focus our discussion based on marginal effects reported in columns A and C.

From Table 3.6, even though we have a negative marginal effect, i.e. lower probability of having a deposit with a bank if household head is female (relative to household headed by male), it is not statistically significant. On the other hand, we find that households headed by female are 6% more likely to have deposits with non-bank formal sources relative to households headed by male. Among the other household characteristics included in the

[22]From 10% more likely to have a bank account if individual is literate and below primary to 119% more likely to have a bank account if the individual is post graduate and above (both relative to an illiterate individual)

[23]i.e. if they have a deposit in other fixed income deposits (such as National Savings Certificate, Kisan Vikas Patra, savings bonds, other small savings schemes), deposits in cooperative banks, NBFCs, co-op credit societies, MFIs, and SHGs.

Table 3.5: Results - Determinants of Individual's access to bank account and debit/credit cards

Outcome of interest	Pr(Individual has a bank account) (Probit)	Pr(Individual has a bank account) (Logit)	Pr(Individual has debit/credit card) (Probit)	Pr(Individual has debit/credit card) (Logit)
	(A)	(B)	(C)	(D)
Dummy = 1 if individual is female	-0.26***	-0.47***	-0.64***	-1.19***
	(0.013)	(0.024)	(0.013)	(0.023)
Dummy = 1 if other household members have account	2.26***	4.01***		
	(0.009)	(0.017)		
Dummy = 1 if other household members have debit or credit			2.10***	3.68***
			(0.008)	(0.015)
Dummy = 1 if individual owns land	0.46***	0.88***	0.22***	0.38***
	(0.015)	(0.030)	(0.012)	(0.021)
Dummy = 1 if other household members own land	-0.10***	-0.18***	-0.14***	-0.21***
	(0.014)	(0.026)	(0.013)	(0.023)
Age of individual (yrs)	-0.00***	-0.00*	-0.02***	-0.03***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of adults in the household	-0.16***	-0.28***	-0.24***	-0.44***
	(0.010)	(0.019)	(0.010)	(0.016)
Dummy =1 if max education in hh is between primary and higher secondary	-0.14***	-0.23***	0.03*	0.05*
	(0.013)	(0.025)	(0.016)	(0.028)
Dummy =1 if max education in hh is grad and above	-0.18***	-0.31***	-0.17***	-0.32***
	(0.017)	(0.032)	(0.018)	(0.033)
Dummy =1 if hh religion is Islam	-0.08***	-0.16***	-0.03***	-0.07***
	(0.012)	(0.022)	(0.010)	(0.019)
Dummy =1 if hh religion is Christian	-0.18***	-0.12***	0.10***	0.11***
	(0.018)	(0.038)	(0.016)	(0.033)
Dummy =1 if hh religion is Others	0.01	0.06	0.04**	0.06*
	(0.021)	(0.041)	(0.018)	(0.035)
Dummy = 1 if hh caste is SC	0.03**	0.03	-0.05***	-0.11***
	(0.012)	(0.022)	(0.010)	(0.019)
Dummy = 1 if hh caste is ST	-0.04***	-0.07***	-0.10***	-0.16***
	(0.014)	(0.026)	(0.012)	(0.022)
Dummy = 1 if hh caste is OBC	-0.01	-0.01	-0.03***	-0.09***
	(0.010)	(0.018)	(0.008)	(0.015)
Dummy = 1 if hh is in Urban area	-0.05***	-0.08***	0.15***	0.31***
	(0.009)	(0.017)	(0.008)	(0.013)
Other individual characteristics	Yes (individual's relation to head of the household, education level of individual, individuals employment)			
Other household characteristics	Yes (asset quartile, monthly expenditure quartiles, age of household head and its square, size of the household, number of children in household)			
State level features	Yes	State FE	Yes	State FE
Observations	285,519	289,512	285,519	289,512
Pseudo R square	0.362	0.366	0.467	0.471

Note: Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

model, we find that probability of having a deposit account in bank increases with the age of the household head. We also observe a U-shaped relationship between probability of having a deposit account in a bank and the quartile to which the household belongs to - in terms both immovable assets as well as monthly expenditure. Relative to households at the bottom-most quartile, the probability of having a deposit with a bank is higher for households in the second quartile and highest for households in the third quartile. However, this probability of having a deposit in a bank is lower for households in the fourth quartile. Households in the upper most quartiles have access to a wider variety of savings instruments. Therefore, they could be holding their wealth in other forms as found by Badarinza et al. (2016). However, that the probability of having a deposit with non-bank formal sources increases with the asset and expenditure quartiles. Households who identify as Muslim have lower probability of having a deposit in banks and in non-formal bank sources relative to households identifying as Hindu. This can be explained by the adherence within Islam to "Riba is haram", i.e, religious prohibition in subscribing to interest-bearing products Narayana and Shagishna (2020). We also observe that households that do not have regular salaried employment are less likely to have deposits with both bank and non-bank formal sources. This is evident from the statistically significant and negative marginal effects for the three dummy variables capturing employment of the household - Dummy = 1 if household identifies as self-employed, Dummy = 1 if household identifies as casual employed, and Dummy = 1 if household identifies as Others (non-regular) for employment.

Table 3.7 reports the marginal effects from Probit estimation of household indicators for access to credit from formal and informal sources, along with credit in-kind[24]. In this analysis, we also control for households deposits with banks and non-bank formal sources. We do so by including a binary variable indicator as well as the amount of deposit held by the household in banks and in non-bank formal sources. This helps control for household's familiarity with

[24]Estimates using Logit regression with state fixed effects are reported in Table C.3 in Appendix B. The estimated marginal effects are qualitatively similar across the two specifications.

Table 3.6: Results - Determinants of household's access to bank and non-bank deposits

Outcome of interest	Household has deposit with Bank (Probit) (A)	Household has deposit with Bank (Logit) (B)	Household has deposit with non-bank formal sources (Probit) (C)	Household has deposit with non-bank formal sources (Logit) (D)
Dummy =1 if hh head is female (0 otherwise)	-0.03	-0.06	0.06***	0.09***
	(0.022)	(0.048)	(0.017)	(0.033)
Age of hh head	0.04***	0.09***	0.02***	0.03***
	(0.003)	(0.006)	(0.003)	(0.005)
Size of household	0.07***	0.16***	-0.01**	0.00
	(0.009)	(0.022)	(0.005)	(0.010)
Dummy =1 if hh religion is Islam	-0.06**	-0.10*	-0.26***	-0.57***
	(0.025)	(0.056)	(0.021)	(0.044)
Dummy =1 if hh religion is Christian	0.02	0.15	-0.24***	-0.34***
	(0.038)	(0.103)	(0.029)	(0.061)
Dummy =1 if hh religion is Others (Sikhism, Jainism, Buddhism, Zoorastrianism)	0.02	-0.06	-0.06*	-0.16**
	(0.046)	(0.110)	(0.031)	(0.068)
Dummy = 1 if hh caste is SC	0.03	0.03	-0.11***	-0.16***
	(0.026)	(0.057)	(0.019)	(0.037)
Dummy = 1 if hh caste is ST	-0.12***	-0.28***	-0.17***	-0.33***
	(0.028)	(0.062)	(0.022)	(0.046)
Dummy = 1 if hh caste is OBC	0.08***	0.13***	-0.09***	-0.05*
	(0.020)	(0.048)	(0.014)	(0.028)
Dummy = 1 if hh identifies as casual employed	-0.15***	-0.36***	-0.13***	-0.27***
	(0.026)	(0.059)	(0.019)	(0.036)
Dummy = 1 if hh identifies as self employed	-0.09***	-0.23***	-0.07***	-0.12***
	(0.025)	(0.057)	(0.015)	(0.030)
Dummy = 1 if hh identifies as Others for employment	-0.25***	-0.51***	-0.06***	-0.13***
	(0.030)	(0.066)	(0.023)	(0.045)
Dummy = 1 if hh is in Urban area	-0.13***	-0.32***	0.07***	0.17***
	(0.018)	(0.041)	(0.013)	(0.026)
Second asset quartile	0.26***	0.62***	0.05***	0.10***
	(0.020)	(0.044)	(0.018)	(0.037)
Third asset quartile	0.39***	0.88***	0.10***	0.22***
	(0.024)	(0.054)	(0.018)	(0.037)
Fourth asset quartile	0.36***	0.82***	0.11***	0.24***
	(0.026)	(0.059)	(0.019)	(0.039)
Second quartile - monthly expenditure	0.17***	0.39***	0.13***	0.21***
	(0.022)	(0.048)	(0.018)	(0.037)
Third quartile - monthly expenditure	0.22***	0.52***	0.18***	0.31***
	(0.025)	(0.058)	(0.019)	(0.038)
Fourth quartile - monthly expenditure	0.10***	0.27***	0.23***	0.38***
	(0.029)	(0.066)	(0.021)	(0.041)
State level controls	Yes	State FE	Yes	State FE
Observations	109,930	110,290	109,930	111,508
Pseudo R square	0.115	0.132	0.0667	0.0898

Note: Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

formal financial resources as well as their credit worthiness. Households with deposits the lender are more credit-worthy than households without any prior relationship with the lender as it reduces information asymmetry in evaluating the credit worthiness of the household.

We find that households headed by a female are 9% less likely to have accessed credit from a formal source relative to a household headed by a male. Further, we find that households with female heads are 15% less likely to have availed credit from banks relative to households headed by males. However, female headed households are 6% more likely to have borrowed from non-bank financial resources, relative to male headed households. This is consistent with the findings in Table 3.6 where female headed households were more likely to have deposits with non-bank formal sources and less likely to have deposits with banks (relative to male headed households). With regard to informal sources of credit, we find that households with female heads are 3% less likely to have availed a loan from such sources relative to households headed by males. We do not find any statistically significant differences in likelihood of borrowing in-kind between households headed by males and females. Households with deposits with bank and non-bank formal sources are more likely to avail credit from formal sources relative to households that do not hold any such deposits with formal sources.

Households are more likely to borrow across different sources as the age of the household head increases. Also, the likelihood of borrowing increases with household's asset quartiles and expenditure quartiles. Swain (2002) also finds evidence that access to formal finance increases with size of land owned. This could be because households in the upper quartiles in terms of asset holding are more likely to be able to meet the collateral requirements for formal loans and informal loans. However, asset and expenditure quartiles are statistically insignificant in determining access to in-kind loans. Households who either have causal employment are less 6% likely to borrow from formal sources and 22% less likely to borrow from banks, relative to households who have regular/salaried employment. Households with casual employment are 20% more likely to borrow from informal sources. We also find

evidence for behavior consistent with Muslim households lack of access for interest-bearing credit products. Looking across columns A to E for "Dummy=1 if hh religion is Islam", we can see that these households are significantly less likely to borrow from formal sources of credit (7% less) and non-bank sources of formal credit (9% less), relative to Hindu households. However, Muslim households are 15% more likely to access in-kind loans relative to Hindu households. Households in urban areas have a preference to use formal sources of credit over informal and in-kind credit. These households are 5% more likely to borrow from formal sources of credit, 21% less likely to borrow from informal sources, and 15% less likely to avail in-kind loans.

We examine the intensity with which different household types use bank deposits, formal credit, informal credit and in-kind loans in Tables 3.8 and 3.9. These two tables present the estimates from the two stages of the Double Hurdle model. Note that we cannot directly interpret these reported coefficients. To interpret the coefficients from the Double Hurdle model, we estimate the Average Partial Effect (APE) by applying the steps outlined in Burke (2009) and use the APE's to draw inferences about the direction and magnitude of the effect of independent variables on the outcome.

Reviewing the estimated coefficients in column B of Table 3.8, we find that conditional on access to deposits with banks (column A), households headed by females hold less deposit in banks relative to households headed by male. This could be due to female headed households being poorer on average than male headed households[25]. However, we have included asset quartiles and expenditure quartiles as control features in these specifications. Therefore, holding all else constant (including household wealth and expenditure), the average partial effect of the household headed by a female relative to a male on use of deposits in banks is -12.6%. That is households with female heads hold 12.6% less deposits with banks relative to deposits held by households with male heads. Age of the head of the household and the

[25]This is based on the evidence discussed in Table 3.2 in Section 3

Table 3.7: Results - Determinant's of household's access to credit from formal sources, informal sources, and in-kind loans

Outcome of interest	Household has access to formal credit	Household has access to formal credit - Banks	Household has access to formal credit - non bank entities	Household has access to informal credit	Household has access to in-kind loans
	(A)	(B)	(C)	(D)	(E)
Dummy =1 if hh head is female (0 otherwise)	-0.09***	-0.15***	0.06***	-0.03**	0.03
	(0.013)	(0.014)	(0.015)	(0.013)	(0.019)
Age of hh head	0.03***	0.03***	0.02***	0.02***	0.01***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Size of household	0.05***	0.02***	0.05***	0.03***	0.00
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Dummy =1 if max educ in hh is between primary and higher secondary	0.11***	0.12***	0.06***	-0.02	0.04**
	(0.015)	(0.017)	(0.018)	(0.014)	(0.021)
Dummy =1 if max educ in hh is grad and above	0.12***	0.23***	-0.12***	-0.22***	-0.05*
	(0.017)	(0.019)	(0.022)	(0.018)	(0.026)
Dummy =1 if hh religion is Islam	-0.07***	-0.01	-0.09***	0.00	0.15***
	(0.013)	(0.014)	(0.016)	(0.014)	(0.018)
Dummy =1 if hh religion is Christian	-0.17***	0.03	-0.33***	0.11***	-0.13***
	(0.020)	(0.022)	(0.029)	(0.021)	(0.037)
Dummy =1 if hh religion is Others (Sikhism, Jainism, Buddhism, Zoorastrianism)	-0.06***	0.05**	-0.15***	-0.06**	0.15***
	(0.022)	(0.024)	(0.029)	(0.024)	(0.033)
Dummy = 1 if hh caste is SC	0.02	-0.11***	0.17***	0.17***	0.10***
	(0.013)	(0.014)	(0.016)	(0.014)	(0.019)
Dummy = 1 if hh caste is ST	-0.09***	-0.18***	0.09***	-0.06***	-0.10***
	(0.015)	(0.017)	(0.019)	(0.016)	(0.024)
Dummy = 1 if hh caste is OBC	-0.06***	-0.09***	0.04***	0.22***	0.08***
	(0.010)	(0.011)	(0.013)	(0.011)	(0.015)
Dummy = 1 if hh identifies as casual employed	-0.06***	-0.22***	0.14***	0.20***	0.07***
	(0.013)	(0.014)	(0.016)	(0.014)	(0.021)
Dummy = 1 if hh identifies as self employed	0.11***	0.08***	0.08***	0.17***	0.11***
	(0.011)	(0.012)	(0.014)	(0.012)	(0.018)
Dummy = 1 if hh identifies as Others for employment	-0.26***	-0.20***	-0.20***	-0.10***	-0.10***
	(0.018)	(0.020)	(0.024)	(0.019)	(0.029)
Dummy = 1 if hh is in Urban area	0.05***	0.03***	0.04***	-0.21***	-0.15***
	(0.009)	(0.010)	(0.012)	(0.010)	(0.014)
Second asset quartile	0.12***	0.17***	0.01	0.04***	-0.01
	(0.012)	(0.014)	(0.014)	(0.012)	(0.017)
Third asset quartile	0.27***	0.38***	-0.01	0.00	0.01
	(0.012)	(0.014)	(0.015)	(0.013)	(0.018)
Fourth asset quartile	0.41***	0.56***	-0.08***	-0.07***	-0.01
	(0.013)	(0.015)	(0.017)	(0.014)	(0.020)
Second quartile - monthly expenditure	0.16***	0.13***	0.11***	0.11***	0.01
	(0.012)	(0.014)	(0.015)	(0.012)	(0.018)
Third quartile - monthly expenditure	0.22***	0.22***	0.11***	0.10***	0.00
	(0.013)	(0.014)	(0.016)	(0.013)	(0.020)
Fourth quartile - monthly expenditure	0.32***	0.35***	0.07***	0.08***	0.02
	(0.015)	(0.016)	(0.019)	(0.015)	(0.022)
Dummy =1 if hh has deposits with bank	0.51***	0.55***	0.27***	0.07***	0.04
	(0.026)	(0.032)	(0.031)	(0.023)	(0.035)
Dummy =1 if hh has deposits with non-bank formal sources	0.13***	0.10***	0.09***	0.09***	0.06**
	(0.015)	(0.015)	(0.018)	(0.015)	(0.023)
State level controls	Yes	Yes	Yes	Yes	Yes
Observations	109,930	109,930	109,930	109,930	109,930
Pseudo R square	0.0687	0.0903	0.0458	0.0526	0.0363

Note: Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

highest education level among members of the household have positive impact on use of deposits held at banks. This is intuitive as savings will be higher in households with older members as they could have accumulated their savings over longer range of time and are more likely to use bank deposits for savings. Similarly, more education within the household empowers household members to better utilize the formal financial instruments available for savings.

Households who do not have regular or salaried employment are use deposits with banks less intensively. Households who identify as self-employed hold 24% less in deposits with banks compared to households with regular or salaried employment while households who are engaged in casual employment hold 46% in deposits with banks. Intuitively, we expect households with more stable income to be able to manage their savings through bank deposit instruments such as recurring deposits or term/fixed deposits. Relative to households belonging to the General caste, we find that households belonging to lower social groups hold less in bank deposits. Scheduled Caste households hold 29% less, while households in Other Backward Classes hold 19% less, and households belonging to Scheduled Tribes hold 7.5% less in deposits with banks relative households belonging to the General caste. The evidence of caste based disparities in access to finance is consistent with the findings of Kumar and Venkatachalam (2019) who use IHDS data on rural farm households from 2011-12. They find that households belonging to scheduled tribes are most likely to face taste-based discrimination when applying for credit. We also find that the amount deposits held in banks increases with household wealth and expenditure quartiles, i.e. richer households are holding more of their savings in bank deposits than poorer households. Households in urban areas also hold greater amounts in deposits with banks compared to households in rural areas, holding all else constant. These findings also calls for development of better financial products for savings to meet the needs of households without regular employment and households from lower socio-economic groups.

Table 3.8: Results - Household's use of bank deposits and credit from formal sources

Outcome of interest	Use of bank deposits: log(amount in bank deposits)		Use of formal credit: log(credit from formal source)			
	First Stage	Second Stage	First Stage	Second Stage	First Stage	Second Stage
	(A)	(B)	(C)	(D)	(E)	(F)
Dummy =1 if hh head is female (0 otherwise)	-0.03	-0.12***	-0.09***	-0.13***	-0.07***	-0.13***
	(0.022)	(0.015)	(0.013)	(0.018)	(0.020)	(0.020)
Median Interest Rate on household's credit loans						-0.01***
						(0.001)
Age of hh head	0.04***	0.02***	0.03***	0.01***	0.01***	0.01**
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Size of household	0.07***	-0.00	0.05***	-0.04***	0.02***	-0.04***
	(0.009)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Dummy =1 if max educ in hh is between primary and higher secondary	0.40***	0.47***	0.11***	0.11***	0.07***	0.10***
	(0.022)	(0.016)	(0.015)	(0.019)	(0.022)	(0.022)
Dummy =1 if max educ in hh is grad and above	0.54***	1.05***	0.12***	0.56***	0.18***	0.54***
	(0.030)	(0.020)	(0.017)	(0.023)	(0.026)	(0.026)
Dummy =1 if hh religion is Islam	-0.06**	-0.02	-0.07***	0.04**	-0.13***	0.03
	(0.025)	(0.016)	(0.013)	(0.018)	(0.020)	(0.021)
Dummy =1 if hh religion is Christian	0.02	0.43***	-0.17***	0.38***	-0.20***	0.32***
	(0.038)	(0.024)	(0.020)	(0.029)	(0.031)	(0.032)
Dummy =1 if hh religion is Others (Sikhism, Jainism, Buddhism, Zoorastrianism)	0.02	0.44***	-0.06***	0.16***	-0.04	0.19***
	(0.046)	(0.031)	(0.022)	(0.033)	(0.035)	(0.036)
Dummy = 1 if hh caste is SC	0.03	-0.35***	0.02	-0.27***	-0.04**	-0.29***
	(0.026)	(0.016)	(0.013)	(0.017)	(0.020)	(0.020)
Dummy = 1 if hh caste is ST	-0.12***	-0.08***	-0.09***	-0.29***	0.00	-0.33***
	(0.028)	(0.018)	(0.015)	(0.020)	(0.024)	(0.023)
Dummy = 1 if hh caste is OBC	0.08***	-0.21***	-0.06***	-0.15***	-0.13***	-0.16***
	(0.020)	(0.013)	(0.010)	(0.013)	(0.015)	(0.015)
Dummy = 1 if hh identifies as casual employed	-0.15***	-0.61***	-0.06***	-0.43***	-0.20***	-0.46***
		(0.015)	(0.013)	(0.018)	(0.020)	(0.020)
Dummy = 1 if hh identifies as self employed	-0.09***	-0.28***	0.11***	-0.17***	0.00	-0.20***
	(0.025)	(0.014)	(0.011)	(0.016)	(0.018)	(0.017)
Dummy = 1 if hh identifies as Others for employment	-0.25***	-0.13***	-0.26***	-0.09***	-0.13***	-0.09***
	(0.030)	(0.022)	(0.018)	(0.028)	(0.030)	(0.031)
Dummy = 1 if hh is in Urban area	-0.13***	0.48***	0.05***	0.56***	0.12***	0.64***
	(0.018)	(0.011)	(0.009)	(0.013)	(0.014)	(0.014)
Second asset quartile	0.26***	0.10***	0.12***	0.03*	0.11***	-0.00
	(0.020)	(0.013)	(0.012)	(0.016)	(0.018)	(0.019)
Third asset quartile	0.39***	0.38***	0.27***	0.27***	0.27***	0.23***
	(0.024)	(0.014)	(0.012)	(0.017)	(0.019)	(0.019)
Fourth asset quartile	0.36***	0.77***	0.41***	0.65***	0.41***	0.62***
	(0.026)	(0.016)	(0.013)	(0.018)	(0.020)	(0.021)
Second quartile - monthly expenditure	0.17***	0.23***	0.16***	0.17***	0.08***	0.19***
	(0.022)	(0.014)	(0.012)	(0.016)	(0.019)	(0.019)
Third quartile - monthly expenditure	0.22***	0.41***	0.22***	0.34***	0.16***	0.35***
	(0.025)	(0.015)	(0.013)	(0.017)	(0.020)	(0.019)
Fourth quartile - monthly expenditure	0.10***	0.69***	0.32***	0.66***	0.26***	0.68***
	(0.029)	(0.018)	(0.015)	(0.019)	(0.022)	(0.022)
Dummy =1 if hh has deposits with bank			0.51***	0.01	0.53***	-0.08
			(0.026)	(0.047)	(0.044)	(0.062)
Dummy =1 if hh has deposits with non-bank formal sources			0.13***	0.04**	0.05**	0.04*
			(0.015)	(0.019)	(0.022)	(0.020)
Other household characteristics	Yes (square of age of household head and number of children)					
State level controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109,930	109,930	109,930	109,930	51,159	51,159
Log Likelihood	-210512	-210512	-138988	-138988	-81605	-81605

Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

Moving on to the use of credit from formal sources, columns C and D in Table 3.8 reports the results from the Double Hurdle model without controlling for the interest rates paid by households on their credit from formal sources. Columns E and F in Table 3.8 reports the estimates from the Double Hurdle model while controlling for the interest rate on credit in the second stage. In both sets of specifications, we control for whether households deposits with any formal sources. The sign for median interest rate is negative and statistically significant, i.e. households borrow less from formal sources when the interest rates are higher. This result is intuitive as the interest rate reflects the price of the credit/loan and we expect an inverse relationship between the two. We also find that the credit availed by households headed by women from formal sources is 29.9% (based on estimate in col F) less compared to credit from formal sources available by households headed by men. The credit amount from formal sources for households belonging scheduled caste is 24.2% less than households in general social group, while households belonging to scheduled tribe have 38% less credit from formal sources and households in other backward classes have 26.4% less, than households in the general social group. Households who are casually employed avail 36% less credit from sources relative to households with regular employment. Households in urban areas avail more credit from formal sources relative to households in rural areas. We also see that the amount of credit from formal sources increases with wealth. This again points towards richer households having greater likelihood of meeting collateral requirements and therefore having greater access and ability to use credit from formal sources. Similar to the results on deposits held in bank, these findings call for better designed financial instruments to improve access and use of formal credit for households without regular employment and households belonging to backward social classes.

Table 3.9 reports results from the Double Hurdle model where the outcome variables are use of informal credit and in-kind loans. Columns A and B report the first and second stage estimates of the Double Hurdle model but does not control for the median interest rate paid by the household on its informal loans. Estimates reported in columns E and F include

median interest rates on informal loans as a control variable. In all specifications reported in Table 3.9, we include features on whether household holds deposits with formal sources (bank and non-bank) as well as the amount of deposits held. We find that size of informal credit availed by households headed by females is 4% less relative to households headed by males.

Households use of informal finance decreases with education. We find that households, where the highest educated member has education between primary and higher secondary, hold 42% less in informal credit compared to households with highest educated member having education up to primary level. Similarly, households where highest educated member has education more than higher secondary hold 74% less in informal credit compared to households with lowest education levels. We also observe a similar difference between households in rural and urban areas where informal credit availed by households in urban areas is 66% less than that held by households in rural areas. Households where the main employment is casual employment, hold 182% more in informal credit compared to households with regular employment. For households where the main employment is self-employment, we find that such households 73% more credit compared to households with regular employment. Households in Scheduled Caste hold 72% more in informal credit and households in Other Backward Castes hold 179% more in informal credit compared to households in general social category. Households belonging to Scheduled Tribes, however, hold 40% less in informal loans relative to households in general social category. We also find that having deposits with bank and non-bank sources positively affect utilization of informal credit.

Table 3.9: Results - Household's use of credit from informal sources and in-kind loans

Outcome of interest	Use of informal credit: log(credit from informal sources)				Use of in-kind loans: log(value of in-kind loans)	
	First Stage (A)	Second Stage (B)	First Stage (C)	Second Stage (D)	First Stage (E)	Second Stage (F)
Dummy =1 if hh head is female (0 otherwise)	-0.03**	-0.05**	0.09***	-0.04	0.03	-0.10**
	(0.013)	(0.022)	(0.020)	(0.025)	(0.019)	(0.050)
Median Interest Rate on household's credit loans				0.01***		
				(0.000)		
Age of hh head	0.02***	0.04***	0.00	0.03***	0.01***	0.03***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.007)
Size of household	0.03***	0.02***	0.00	0.02***	0.00	0.03**
	(0.004)	(0.006)	(0.005)	(0.007)	(0.005)	(0.015)
Dummy =1 if max educ in hh is between primary and higher secondary	-0.02	0.07***	-0.11***	0.08***	0.04**	0.13**
	(0.014)	(0.023)	(0.023)	(0.026)	(0.021)	(0.052)
Dummy =1 if max educ in hh is grad and above	-0.22***	0.27***	-0.32***	0.29***	-0.05*	0.35***
	(0.018)	(0.029)	(0.026)	(0.033)	(0.026)	(0.066)
Dummy =1 if hh religion is Islam	0.00	0.11***	0.08***	0.16***	0.15***	0.16***
	(0.014)	(0.022)	(0.020)	(0.025)	(0.018)	(0.049)
Dummy =1 if hh religion is Christian	0.11***	0.21***	0.17***	0.27***	-0.13***	0.21**
	(0.021)	(0.035)	(0.030)	(0.040)	(0.037)	(0.107)
Dummy =1 if hh religion is Others (Sikhism, Jainism, Buddhism, Zoroastrianism)	-0.06**	0.05	-0.04	0.12***	0.15***	0.36***
	(0.024)	(0.040)	(0.034)	(0.045)	(0.033)	(0.093)
Dummy = 1 if hh caste is SC	0.17***	-0.16***	0.12***	-0.20***	0.10***	-0.14***
	(0.014)	(0.023)	(0.019)	(0.025)	(0.019)	(0.051)
Dummy = 1 if hh caste is ST	-0.06***	-0.61***	-0.04*	-0.61***	-0.10***	-0.12**
	(0.016)	(0.027)	(0.023)	(0.031)	(0.024)	(0.061)
Dummy = 1 if hh caste is OBC	0.22***	0.02	0.24***	0.00	0.08***	-0.09**
	(0.011)	(0.019)	(0.015)	(0.021)	(0.015)	(0.042)
Dummy = 1 if hh identifies as casual employed	0.20***	-0.04*	0.24***	-0.07***	0.07***	-0.13**
	(0.014)	(0.023)	(0.020)	(0.026)	(0.021)	(0.054)
Dummy = 1 if hh identifies as self employed	0.17***	0.06**	0.13***	0.04*	0.11***	0.08
	(0.012)	(0.022)	(0.017)	(0.024)	(0.018)	(0.050)
Dummy = 1 if hh identifies as Others for employment	-0.10***	-0.02	0.14***	0.07*	-0.10***	-0.09
	(0.019)	(0.038)	(0.029)	(0.042)	(0.029)	(0.084)
Dummy = 1 if hh is in Urban area	-0.21***	0.23***	-0.31***	0.30***	-0.15***	0.14***
	(0.010)	(0.017)	(0.014)	(0.019)	(0.014)	(0.041)
Second asset quartile	0.04***	0.11***	-0.05**	0.09***	-0.01	0.01
	(0.012)	(0.019)	(0.019)	(0.022)	(0.017)	(0.044)
Third asset quartile	0.00	0.28***	-0.15***	0.26***	0.01	0.15***
	(0.013)	(0.021)	(0.019)	(0.023)	(0.018)	(0.047)
Fourth asset quartile	-0.07***	0.51***	-0.27***	0.49***	-0.01	0.30***
	(0.014)	(0.023)	(0.020)	(0.026)	(0.020)	(0.052)
Second quartile - monthly expenditure	0.11***	0.26***	0.01	0.21***	0.01	0.11**
	(0.012)	(0.021)	(0.019)	(0.023)	(0.018)	(0.046)
Third quartile - monthly expenditure	0.10***	0.38***	-0.02	0.33***	0.00	0.26***
	(0.013)	(0.022)	(0.019)	(0.025)	(0.020)	(0.050)
Fourth quartile - monthly expenditure	0.08***	0.63***	-0.07***	0.56***	0.02	0.47***
	(0.015)	(0.025)	(0.021)	(0.028)	(0.022)	(0.058)
Dummy =1 if hh has deposits with bank	0.07***	0.25***	-0.39***	0.20***	0.04	0.08
	(0.023)	(0.042)	(0.046)	(0.053)	(0.035)	(0.095)
Dummy =1 if hh has deposits with non-bank formal sources	0.09***	0.09***	0.04*	0.12***	0.06**	0.13**
	(0.015)	(0.026)	(0.021)	(0.029)	(0.023)	(0.061)
Other household characteristics	Yes (square of age of household head, employment, number of children)					
State level controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109,930	109,930	51,159	51,159	109,930	109,930
Log Likelihood	-120156	-120156	-72553	-72553	-39319	-39319

Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001

Estimating the average predictive effect for in-kind loans, we find that households headed by female hold 26% more in-kind loans compared to male headed households. Households belonging to Scheduled Caste have 10.4% more, while households belonging to Scheduled Tribe have 10.7% less, and households belonging to Other Backward Castes have 8.2% more in in-kind loans relative to households in general social category. In terms of employment, households with casual employment hold 6.1% more while households with self-employment hold 12.5% more in in-kind loans than households with regular employment. In households where the highest educated member has more than higher secondary education, their in-kind loans are 2.7% less than households where the highest educated member has less than primary education. We also find that households belonging to Islam religion hold 18.2% more in in-kind loans compared to households belonging to Hindu religion. Households in urban areas have 13.4% less in-kind loans relative to households in rural areas. These findings make intuitive sense because if these households (those with casual and self employment, those belonging to SC and OBC social strata, those belonging to Muslim religion, and those headed by women headed households) are unable to meet the need for credit from formal sources, then their remaining choices are informal sources and borrowing in-kind.

3.6 Conclusion

On summing up the empirical evidence, we conclude that there exists disparities across different dimensions in access and use of financial resources by individuals and households. There is evidence indicating that women and households headed by women have less access to formal financial savings and credit instruments. Putting together the estimates from models on access to formal savings instruments, formal and informal credit, and in-kind loans, we find households headed by women are more likely to tap into the informal market, including in-kind loans to meet their requirement for credit. At the individual level, we find evidence

that women are less likely than men to have a bank account and a debit or credit card. We also observe that the households in rural areas with lower levels of education, casual and self employed, and belong to lower social strata use informal cash credit and in-kind loans more intensively. These findings in conjunction with the existing findings from Ghosh and Dharmarjan (2017) implies that the gender disparities in access to finance have not improved in past decade despite the many schemes undertaken by the Ministry of Women and Child Development, Government of India [26].

There are two limitations to the current study. First, I am unable to build a structural model of demand and supply to evaluate access to finance and its usage due to limitations in the data available. The data available is observational in nature and only collects information where households have been successful in accessing credit from any source. It does not collect any information on cases where the households applied for credit but did not receive it. Second, though mobile money and digital transactions are becoming increasingly important as a financial service, we do not have data in the AIDIS survey to study potential gender-based discrepancies within such transactions.

However, despite these limitations, I find evidence of gender disparities in access to finance and its use from both formal and informal sources. This is particularly important in the context of the ongoing efforts and initiatives to increase account ownership and use of digital payments by the Indian government [27]. FINDEX 2021 found that about two-thirds of adults who were unbanked could not use an account at a financial institution without help (if they opened one). This is a pertinent issue as financially inexperienced users may not understand and be able to use financial services to optimize its benefits and minimize risks (Demirguc-Kunt et al., 2022). FINDEX 2021 also found that women are 5 percentage points more likely than men to need help in using their mobile money account. However, inexperienced

[26]The list of schemes for women empowerment can be found here - `https://wcd.nic.in/schemes-listing/2405`

[27]List of current schemes by Ministry of Electronics and Information Technology to promote digital payments are available here - `https://www.meity.gov.in/incentivepromotional-schemes`

users who are dependent on other family members to use their financial account can be more vulnerable and at a higher risk of financial abuse. In India, a Gallup poll from 2021 finds that while 80% of men have access to a mobile device, only 60% of women does. The findings in this study also highlight the limitations of the Jan Dhan Yojana programme. Though the programme succeeded in increasing number of savings account opened, it has not led to increase in use of such accounts. There still exists barriers that prevent women and members of the lower social-economic strata from effectively accessing financial services. Given these findings, it is pertinent to build in additional measures to improve financial literacy among individuals, especially women to improve their independence. This requires investments in improving numeracy and financial literacy skills, designing financial products based on customer's usage patterns and their capabilities. Policy measures should also be implemented to establish strong consumer safeguards to ensure that consumers have a net benefit from the financial system and have greater trust in the financial system.

Bibliography

Allen, F., Demirguc-Kunt, A., Klapper, L., and Soledad, M. (2016). "The foundations of financial inclusion: Understanding ownership and use of formal accounts." *Journal of Financial Intermediation*, *27*, 1–30.

Aportela, F. (1999). "Effects of financial access on savings by low-income people." *MIT Department of Economics*.

Ashraf, N., Karlan, D., and Yin, W. (2006). "Tying Odysseus to the mast: Evidence from a commitment savings product in the Phillippines." *Quartery Journal of Economics*, *121*(2), 673–697.

Ashraf, N., Karlan, D., and Yin, W. (2010). "Female empowerment: Further evidence from a commitment savings product in the Phillipppines." *World Development*, *28*(3), 333–344.

Badarinza, C., Balasubramaniam, V., and Ramadorai, T. (2016). "The Indian household finance landscape." *SSRN*.

Banerjee, A., and Duflo, E. (2014). "Do firms want to borrow more? testing credit constraints using a directed lending program." *The Review of Economic Studies*, *81(2)*, 572–607.

Basavaraj, A., and Bhattacharjee, M. (2014). "A study of household borrowing in the villages of Channarayapatna block, Hassan district." *Journal of Business and Economic Policy*, *1*, 71–83.

Berger, A., and Mester, L. (1997). "Inside the black box: What explains differences in the efficiencies of financial institutions?" *Journal of Banking and Finance*, *21*(7), 895–947.

Bernheim, B. D., and Whinston, M. (1990). "Multimarket contact and collusive behavior." *RAND Journal of Economics*, *21*(1), 1–26.

Bhattacharyya, A., Lovell, C. A. K., and Sahay, P. (1997). "The impact of liberalization on the productive efficiency of Indian commercial banks." *European Journal of Operational Research*, *98*(2), 332–345.

Bikker, J. A., and Haaf, K. (2002). "Competition, concentration and their relationship: An empirical analysis of the banking industry." *Journal of Banking and Finance*, *26*(11), 2191–2214.

Bresnahan, T. (1982). "The oligopoly solution concept is identified." *Economics Letters*, *10*(1-2), 87–92.

Burgess, R., and Pande, R. (2005a). "Do rural banks matter? evidence from the indian social banking experiment." *American Economic Review*, *95*(3), 780–795.

Burgess, R., and Pande, R. (2005b). "Do rural banks matter? Evidence from the Indian social banking experiment." *American Economic Review*, *95*, 780–795.

Burke, W. (2009). "Fitting and interpreting cragg's tobit alternative using Stata." *Stata Journal*, *9*(4), 584–592.

Campero, A., and Kaiser, K. (2013). "Access to credit: Awareness and use of formal and informal credit institutions." *Working Papers 2013-07*.

Casu, B., Ferrari, A., and Zhao, T. (2013). "Regulatory reform and productivity change in indian banking." *The Review of Economics and Statistics*, *95*(3), 1066–1077.

Coble, D. (2019). "Multimarket contact in banking competition in the united states."

Coccorese, P. (2009). "Market power in local banking monopolies." *Journal of Banking and Finance*, *33*(7), 1196–1210.

Coccorese, P., and Pellecchia, A. (2009). "Multimarket contact and profitability in banking: Evidence from italy." *Journal of Financial Services Research*, *35*(3), 245–271.

Coccorese, P., and Pellecchia, A. (2013). "Multimarket contact, competition and pricing in banking." *Journal of International Money and Finance*, *37*(C), 187–214.

Cragg, J. (1971). "Some statistical models for limited dependent variables with application to the demand for durable goods." *Econometrica*, *39*, 829–844.

Das, A., and Kumbhakar, S. (2016). "Markup and efficiency of indian banks: an input distance function approach." *Empirical Economics*, *51*(4), 1689–1719.

Degryse, H., and Ongena, S. (2005). "Distance, lending relationships, and competition." *Journal of Finance*, *60*(1), 231–266.

Demirguc-Kunt, A., Klapper, L., and Randall, D. (2014). "Islamic finance and financial inclusion: Measuring use of and demand for formal financial services among muslim adults." *Review of Middle East Economics and Finance*, *10(2)*, 177–218.

Demirguc-Kunt, A., Klapper, L., and Singer, D. (2013). "Financial inclusion and legal discrimination against women: Evidence from developing countries." (6416).

Demirguc-Kunt, A., Klapper, L., Singer, D., and Ansar, S. (2022). "The global findex database 2021: Financial Inclusion, digital payments, and resilience in the age of covid-19."

Demirguc-Kunt, A., Klapper, L., Singer, D., and Van Oudheusden, P. (2015). *The Global Findex Database 2014: Measuring Financial Inclusion around the World.* 7255.

Dharmarajan, V. (2023). "Measuring competition in the indian banking industry." *Unpublished manuscript, University of Washington*.

Doepke, M., and Tertilt, M. (2019). "Does female empowerment promote economic development?" *Journal of Economic Growth*, *24*, 309–343.

Dupas, P., and Robinson, J. (2013). "Savings constraints and microenterprise development: Evidence from a field experiment in Kenya." *American Economic Journal: Applied Economics*, *5 (1)*, 163–92.

Evans, W., and Kessides, I. (1994). "Living by the "golden rule": Multimarket contact in the u. s. airline industry." *The Quarterly Journal of Economics*, *109*(2), 341–366.

Feinberg, R. (1984). "Mutual forbearance as an extension of oligopoly theory." *Journal of Economics and Business*, *36*(2), 243–249.

Field, E., Pande, R., Rigo, N., Schaner, S., and Troyer Moore, C. (2021). "On her own account: How strengthening women's financial control impacts labor supply and gender norms." *American Economic Review*, *11 (7)*, 2342–75.

Fletschner, D. (2008). "Women's access to credit: Does it matter for household efficiency?" *American Journal of Agricultural Economics*, *90*, 669–683.

Fuentelsaz, L., and Gomez, J. (2006). "Multipoint competition, strategic similarity and entry into geographic markets." *Strategic Management Journal*, *27*(5), 477–499.

Ghosh, S., and Dharmarjan, V. (2017). "What constrains financial inclusion for women? evidence from Indian micro data." *World Development*, *92*, 60–81.

Girma, S., and Shortland, A. (2008). "The political economy of financial development." *Oxford Economics Paper*, *60*, 567–96.

Govindapuram, S., Bhupatiraju, S., and Sirohi, R. (2023). "Determinants of women's financial inclusion: Evidence from India." *Annals of Public and Cooperative Economics*, *94*(1), 131–158.

Gruben, W., and McComb, R. P. (2003). "Privatization, competition, and supercompetition in the mexican commercial banking system." *Journal of Banking and Finance*, *27*(2), 229–249.

Gulati, R., and Kumar, S. (2016). "Assessing the impact of the global financial crisis on the profit efficiency of indian banks." *Economic Modelling*, *58*, 167–181.

Haveman, H. A., and Nonnemaker, L. (2000). "Competition in multiple geographic markets: The impact on growth and market entry." *Administrative Science Quarterly*, *45*(2), 232–267.

Heggestad, A. A., and Rhoades, S. A. (1978). "Multi-market interdependence and local market competition in banking." *The Review of Economics and Statistics*, *60*(4), 523–32.

Jack, W., and Suri, T. (2014). "Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution." *American Economic Review*, *104*(1), 183–223.

Jans, I., and Rosenbaum, D. I. (1997). "Multimarket contact and pricing evidence from the u.s. cement industry." *International Journal of Industrial Organization*, *15*(3), 391–412.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). "Getting to the top of mind: How reminders increase saving." *Management Science*, *62*(12), 3393–3411.

Kast, F., Meier, S., and Pomeranz, D. (2012). "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device." *National Bureau of Economic Research, Inc.*

Klapper, L., Laeven, L., and Rajan, R. (2006). "Entry regulation as a barrier to entrepreneurship." *Journal of Financial Economics*, *82(3)*, 591–629.

Klasen, S., and Lamanna, F. (2009). "The impact of gender inequality in education and employment on economic growth: New evidence for a panel of countries." *Feminist Economics*, *15*(3), 91–132.

Kumar, A., Pal, R., and Pal, R. (2019). "Usage of formal financial services in India: Demand barriers or supply constraints?" *Economics Modelling*, *80*, 244–259.

Kumar, S. M., and Venkatachalam, R. (2019). "Caste and credit: A woeful tale." *The Journal of Development Studies*, *55*(8), 1816–1833.

Kumbhakar, S., and Sarkar, S. (2003). "Deregulation, ownership, and productivity growth in the banking industry: Evidence from india." *Journal of Money, Credit and Banking*, *35*(3), 403–24.

Kumbhakar, S., and Wang, H.-J. (2006). "Pitfalls in the estimation of a cost function that ignores allocative inefficiency: A monte carlo analysis." *Journal of Econometrics*, *134*(2), 317–340.

Lau, L. J. (1982). "On identifying the degree of competitiveness from industry price and output data." *Economics Letters*, *10*(1-2), 93–99.

Lee, J. N., Morduch, J., Ravindran, S., Shonchoy, A., and Zaman, H. (2021). "Poverty and migration in the digital age: Experimental evidence on mobile banking in Bangladesh." *American Economic Journal: Applied Economics*, *13*(1), 38–71.

McKenzie, D., and Woodruff, C. (2008). "Experimental evidence on returns to capital and access to finance in Mexico." *World Bank Economic Review*, *22(3)*, 457–82.

Mester, L. (1987). "A multiproduct cost study of savings and loans." *Journal of Finance*, *42*(2), 423–45.

Molnar, J., Violi, R., and Zhou, X. (2013). "Multimarket contact in italian retail banking: Competition and welfare." *International Journal of Industrial Organization*, *31*(5), 368–381.

Moore, D., Niazi, Z., Rouse, R., and Kramer, B. (2019). "Building resilience through financialinclusion: A review of existing evidence and knowledge gaps."

Narayana, D., and Shagishna, K. (2020). "The role of religious faith in financial exclusion: An analysis of financial deepening in India." *Indian Public Policy Review*, *2*(1), 45–64.

Pal, D. (2022). "Formal financial outreach in rural India." *Economics and Political Weekly*, *LVII*(29), 101–106.

Perera, S., Skully, M., and Wickramanayake, J. (2006). "Competition and structure of south asian banking: a revenue behaviour approach." *Applied Financial Economics*, *16*(11), 789–801.

Pilloff, S. (1999). "Multimarket contact in banking." *Review of Industrial Organization*, *14*(2), 163–182.

Prasad, A., and Ghosh, S. (2007). "Competition in indian banking." *South Asia Economic Journal*, *8*(2), 265–284.

Prina, S. (2015). "Banking the poor via savings accounts: Evidence from a field experiment." *Journal of Development Economics*, *115*, 16–31.

Rakshit, B., and Bardhan, S. (2019). "Bank competition and its determinants: Evidence from indian banking." *International Journal of the Economics of Business*, *26*(2), 283–313.

Rampal, P., and Biswas, S. (2022). "Socioeconomic determinants of household investment portfolio in India." *Journal of Emerging Market Finance*, *21*, 265–290.

Rhoades, S. A., and Heggestad, A. A. (1985). "Multimarket interdependence and performance in banking: Two tests." *The Antitrust Bulletin*, *30*(4), 975–995.

Schaner, S. (2018). "The persistent power of behavioral change: Long-run impacts of temporary savings subsidies for the poor." *American Economic Journal: Applied Economics*, *10*(3), 67–100.

Sengupta, R., and Vardhan, H. (2020). "Are more productive banks always better?" *Indira Gandhi Institute of Development Research, Mumbai Working Papers*.

Sensarma, R. (2008). "Deregulation, ownership and profit performance of banks: evidence from india." *Applied Financial Economics*, *18*(19), 1581–1595.

Shaffer, S. (2004). "Patterns of competition in banking." *Journal of Economics and Business*, *56*(4), 287–313.

Solomon, E. H. (1970). "Bank merger policy and problems a linkage theory of oligopoly." *Journal of Money, Credit and Banking*, *2*(3), 323–36.

Spagnolo, G. (1999). "On interdependent supergames: Multimarket contact, concavity, and collusion." *Journal of Economic Theory*, *89*(1), 127–139.

Swain, R. (2002). "Credit rationing in rural India." *Journal of Economic Development*, *27(2)*, 1–20.

Thomas, C., and Willig, R. D. (2006). "The risk of contagion from multimarket contact." *International Journal of Industrial Organization*, *24*(6), 1157–1184.

Van Leuvensteijn, M., Bikker, J., Rixtel, A., and Kok, C. (2011). "A new approach to measuring competition in the loan markets of the euro area." *Applied Economics*, *43*(23), 3155–3167.

Whalen, G. W. (1996). "Nonlocal concentration, multimarket linkages, and interstate banking." *The Antitrust Bulletin*, *41*(2), 365–397.

Young, N. (2021). "Banking and growth: Evidence from a regression discontinuity analysis." *Working paper*.

Zhao, T., Casu, B., and Ferrari, A. (2010). "The impact of regulatory reforms on cost structure, ownership and competition in indian banking." *Journal of Banking and Finance*, *34*(1), 246–254.

Appendix A

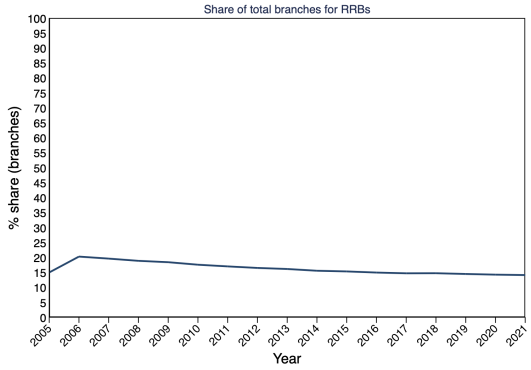
Appendix for Chapter 1

A.1 Background of Banking Industry

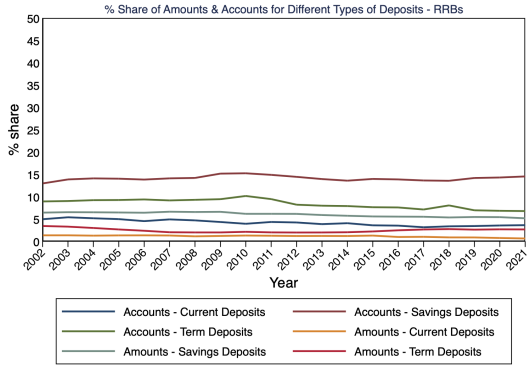
In this appendix, I describe the key events and characteristics of the banking industry in India during the period of study, 2004-2021 to supplement the discussion in the main text.

I do not consider key restriction determining the banks under study is the availability of bank-wise lending rates. This rules out scheduled cooperative banks, regional rural banks, small finance banks, and payments banks from the study. As Figures A.1a-A.1c show, Regional Rural Banks are a small part of the deposit and credit market. They account for 16.3% of all branches in India, 3.28% of all deposits and 2.68% of all credit disbursed in the banking industry. On average, they account for 15.72% of small credit advances in the banking market and 6% of savings account deposits in the banking system during the sample period. Therefore, RRBs are not a significant size of the banking market and can be ignored. I also exclude Small Finance Banks and Payments Banks because they are relatively new entities in the market place and their scope of activities differ significantly from scheduled commercial banks with universal banking licenses (see Table 1.1 Both types of banks started operations in 2016. Small Finance Banks has an average market share per year of 0.3%, 0.7% and 3.6% of total deposits, total credit, and total credit to small borrowers during 2016-2021. Payments Banks on the other hand do not have authority to make credit available to their clients.

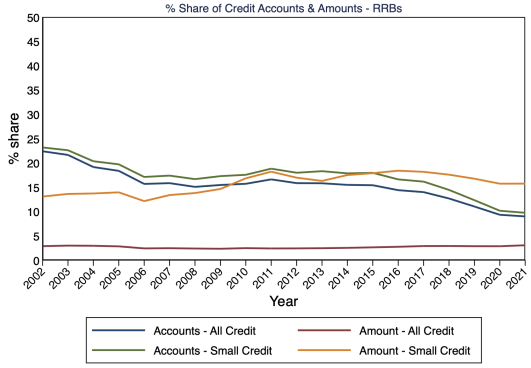
Figure A.1: Statistics on Regional Rural Banks



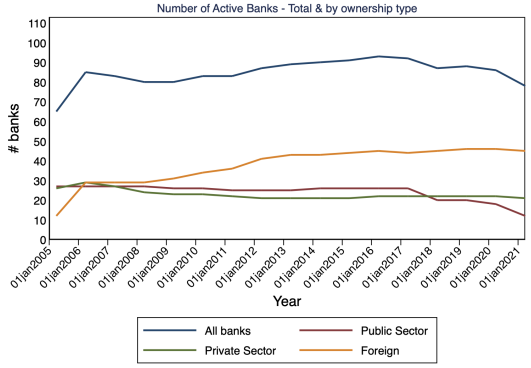
(a) Fig A.1: % share of branches - RRBs, 2005-2021



(b) Fig A.2: % share for each type of deposits (by accounts & amount) - RRBs, 2002-2021



(c) Fig A.3: % share for All Credit & Small Borrowers (by accounts & amount) - RRBs, 2002-2021



(d) Fig A.4: Number of active banks between 2005-2021

To compare banks with similar characteristics and scope of business, we focus on public sector banks, private sector banks, and foreign banks. Figure A.1d summarizes the number of active banks[1] each year by type during the period of study. There are 65 banks altogether at the start of the sample period (2005) which increases to a maximum of 93 banks in 2016 before declining to 78 by 2021. There is a steady decrease in number of Public sector banks as we start with 27 such banks in 2005 and end with only 12 public sector banks by 2021[2].

[1]A bank is defined as active if it has atleast one functional office in the country during the year.

[2]Bharatiya Mahila Bank is the only new public sector bank during the sample period, starting its operations in March 2014. However, it was later amalgamated into State Bank of India in April 2017.

In 2008 and 2010, a public sector bank each was amalgamated with State Bank of India. In April 2017, 6 public sector banks were amalgamated with State Bank of India. In 2019, 2 public sector banks were amalgamated with Bank of Baroda while in 2020, there were three different sets of amalgamations between public sector banks[3]. Thus, the number of public sector banks has decreased during the sample period largely due to the result amalgamations of these banks with their larger counterparts.

Similarly, there are 26 private banks at the start of the sample period in 2005 and this reduces to 21 private banks by 2021. RBI issued universal banking licenses for two private banks at the start of the sample period. Kotak Mahindra Bank started operations in 2003 while Yes Bank started its operations in 2004. There were no new universal banking licenses issued to domestic banks till the creation of Bharatiya Mahila Bank in 2014. In 2015, RBI issued new universal banking licenses to 2 private banks - Bandhan Bank and IDFC First Bank Limited, both of whom started their operations in 2015. Between 2002 and 2021, there were 12 mergers where a private bank merged with another bank. In case of foreign banks, as Figure A.4 shows, the number of foreign banks actively operating in India has increased steadily from 12 foreign banks in 2005 to 45 such banks in 2021. Between 2005 and 2021, 17 foreign banks closed their operations in India while 24 new foreign banks started their operations in India.

[3]Dena Bank and Vijaya Bank were amalgamated with Bank of Baroda in 2019. In 2020, Allahabad Bank was amalgamated with Indian Bank, Syndicate bank with Canara Bank, while United Bank of India and Oriental Bank of Commerce were amalgamated with Punjab National Bank, Corporation Bank and Andhra Bank were amalgamated with Union Bank of India

Appendix B

Appendix for Chapter 2

B.1 Robustness

Table B.1: MMC 1 - Bank level - Including only public and private sector banks

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 1	0.0754*	0.0602*	0.0603*	0.0712*	0.0660**	0.0653*
	(0.0422)	(0.0356)	(0.0339)	(0.0377)	(0.0316)	(0.0360)
NPA/Total Loans	4.340**	5.019**	4.929**	4.631**	4.696**	4.904**
	(1.930)	(2.015)	(2.015)	(1.973)	(1.825)	(1.939)
Total Assets per Branch	-0.0399	-0.0605	-0.0592	-0.0581	-0.0489	-0.0697
	(0.0488)	(0.0453)	(0.0450)	(0.0453)	(0.0439)	(0.0460)
Total Loans/Total Assets	1.791***	1.641**	1.682**	1.548**	1.373**	1.533**
	(0.640)	(0.656)	(0.665)	(0.662)	(0.640)	(0.649)
Equity/Total Assets	-2.418*	-1.866	-2.099	-2.534	-2.384	-2.381
	(1.345)	(1.667)	(1.678)	(1.561)	(1.580)	(1.661)
ln(Branches)	0.516***	0.442***	0.426***	0.434***	0.496***	0.463***
	(0.144)	(0.123)	(0.121)	(0.123)	(0.124)	(0.121)
Operating Expenses/Total Assets	87.56***	89.43***	90.94***	89.32***	84.18***	86.03***
	(7.888)	(8.104)	(7.792)	(8.195)	(8.733)	(8.634)
HHI	5.071	-1.627	-2.982	-0.266	-2.251	6.622
	(7.877)	(7.673)	(7.572)	(8.880)	(8.727)	(8.755)
Observations	797	797	797	797	797	797
Within R-Squared	0.634	0.658	0.668	0.664	0.650	0.669
Between R-Squared	0.0944	0.129	0.138	0.119	0.114	0.150
Overall R-Squared	0.0815	0.122	0.136	0.143	0.148	0.179
Adjusted R-Squared	0.622	0.643	0.649	0.645	0.634	0.649
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

Table B.2: MMC 2 - Bank level - Including only public and private sector banks

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 2	0.0778*	0.0633*	0.0638*	0.0742*	0.0683**	0.0686*
	(0.0419)	(0.0355)	(0.0339)	(0.0376)	(0.0315)	(0.0360)
NPA/Total Loans	4.288**	4.979**	4.889**	4.581**	4.653**	4.855**
	(1.923)	(2.009)	(2.009)	(1.965)	(1.816)	(1.930)
Total Assets per Branch	-0.0389	-0.0593	-0.0579	-0.0568	-0.0475	-0.0685
	(0.0487)	(0.0452)	(0.0449)	(0.0451)	(0.0438)	(0.0459)
Total Loans/Total Assets	1.799***	1.648**	1.689**	1.555**	1.380**	1.539**
	(0.641)	(0.656)	(0.665)	(0.663)	(0.641)	(0.649)
Equity/Total Assets	-2.393*	-1.839	-2.070	-2.503	-2.355	-2.346
	(1.342)	(1.667)	(1.680)	(1.563)	(1.580)	(1.666)
ln(Branches)	0.521***	0.445***	0.428***	0.439***	0.503***	0.467***
	(0.140)	(0.119)	(0.118)	(0.119)	(0.122)	(0.117)
Operating Expenses/Total Assets	87.62***	89.40***	90.89***	89.32***	84.19***	86.02***
	(7.909)	(8.119)	(7.804)	(8.200)	(8.763)	(8.639)
HHI	4.950	-1.846	-3.263	-0.491	-2.498	6.409
	(7.840)	(7.635)	(7.470)	(8.813)	(8.667)	(8.672)
Observations	797	797	797	797	797	797
Within R-Squared	0.634	0.658	0.669	0.665	0.650	0.669
Between R-Squared	0.0989	0.133	0.142	0.123	0.118	0.154
Overall R-Squared	0.0863	0.126	0.141	0.149	0.153	0.184
Adjusted R-Squared	0.623	0.643	0.649	0.645	0.635	0.649
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

Table B.3: MMC 3 - Bank level - Including only public and private sector banks

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 3	1.581*	1.407*	1.468*	1.409*	1.444**	1.270
	(0.933)	(0.803)	(0.761)	(0.800)	(0.691)	(0.785)
NPA/Total Loans	4.669**	5.247**	5.156**	4.986**	4.958**	5.238**
	(1.945)	(2.014)	(2.017)	(2.014)	(1.867)	(1.989)
Total Assets per Branch	-0.0424	-0.0657	-0.0644	-0.0625	-0.0534	-0.0741*
	(0.0481)	(0.0437)	(0.0435)	(0.0427)	(0.0419)	(0.0434)
Total Loans/Total Assets	1.840***	1.685**	1.726**	1.628**	1.425**	1.605**
	(0.660)	(0.664)	(0.672)	(0.687)	(0.653)	(0.670)
Equity/Total Assets	-2.999**	-2.182	-2.413	-3.047**	-2.822*	-2.848*
	(1.314)	(1.544)	(1.566)	(1.416)	(1.488)	(1.506)
ln(Branches)	0.573***	0.463***	0.442***	0.477***	0.530***	0.505***
	(0.143)	(0.119)	(0.117)	(0.116)	(0.126)	(0.114)
Operating Expenses/Total Assets	87.53***	90.01***	91.45***	89.78***	84.45***	86.35***
	(8.060)	(8.179)	(7.903)	(8.363)	(8.776)	(8.704)
HHI	5.870	-0.970	-2.686	-0.416	-1.278	6.766
	(7.597)	(7.287)	(7.154)	(8.794)	(8.526)	(8.445)
Observations	797	797	797	797	797	797
Within R-Squared	0.631	0.657	0.668	0.662	0.648	0.666
Between R-Squared	0.0941	0.136	0.146	0.129	0.122	0.159
Overall R-Squared	0.0871	0.129	0.145	0.157	0.167	0.191
Adjusted R-Squared	0.619	0.642	0.648	0.642	0.633	0.646
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
State variables (levels)	No	Yes	Yes	No	No	No
State variables (growth rates)	No	No	Yes	Yes	No	Yes
State variables (ln(levels))	No	No	No	Yes	No	Yes
State variables (per capita terms)	No	No	No	No	Yes	No
State population	No	No	No	No	No	Yes

B.2 District level analysis

In the main analysis, I defined markets at the state level. In this appendix, I present the results for bank-level models where the multi-market contact measure is computed with markets defined at the district level. I have assumed that banks that were merged or amalgamated into another bank and banks that acquired them were one entity for the entire sample period for this analysis.

Table B.4: District wise MMC 1 - Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 1	0.212***	0.197***	0.184***	0.230***	0.202***	0.232***
	(0.0664)	(0.0684)	(0.0626)	(0.0738)	(0.0611)	(0.0744)
NPA/Total Loans	-1.541	-0.961	-1.058	-1.163	0.338	-1.227
	(2.175)	(2.093)	(2.097)	(2.201)	(1.988)	(2.213)
Total Assets per Branch	0.00110	0.00444	0.00386	0.00273	0.00415	0.00290
	(0.00602)	(0.00806)	(0.00834)	(0.00793)	(0.00848)	(0.00761)
Total Loans/Total Assets	0.788	0.566	0.608	0.707	0.990	0.724
	(1.087)	(0.988)	(0.991)	(0.983)	(0.773)	(0.967)
Equity/Total Assets	2.855	2.755	2.638	2.477	2.802	2.513
	(1.884)	(1.714)	(1.734)	(1.726)	(1.782)	(1.720)
ln(Branches)	0.588**	0.603***	0.627***	0.622***	0.587**	0.576**
	(0.233)	(0.168)	(0.169)	(0.201)	(0.238)	(0.231)
Operating Expenses/Total Assets	60.99***	60.86***	61.61***	60.85***	61.96***	59.56***
	(12.66)	(11.70)	(11.41)	(12.27)	(12.08)	(12.18)
HHI	-21.80*	-23.77**	-22.71*	-8.156	-24.51**	-19.01*
	(11.40)	(9.940)	(12.29)	(12.94)	(10.65)	(11.08)
Observations	1312	1312	1312	1312	1312	1312
Within R-Squared	0.284	0.317	0.323	0.323	0.331	0.325
Between R-Squared	0.239	0.260	0.261	0.229	0.244	0.220
Overall R-Squared	0.195	0.214	0.212	0.186	0.207	0.182
Adjusted R-Squared	0.271	0.300	0.300	0.300	0.314	0.301

Table B.5: District wise MMC 2 - Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 2	0.211***	0.201***	0.187***	0.232***	0.204***	0.235***
	(0.0680)	(0.0710)	(0.0653)	(0.0759)	(0.0623)	(0.0765)
NPA/Total Loans	-1.564	-0.973	-1.071	-1.189	0.326	-1.253
	(2.174)	(2.089)	(2.093)	(2.197)	(1.985)	(2.209)
Total Assets per Branch	0.00110	0.00441	0.00383	0.00275	0.00414	0.00291
	(0.00606)	(0.00810)	(0.00837)	(0.00797)	(0.00852)	(0.00765)
Total Loans/Total Assets	0.792	0.571	0.613	0.710	0.993	0.727
	(1.087)	(0.986)	(0.989)	(0.982)	(0.773)	(0.966)
Equity/Total Assets	2.860	2.766	2.649	2.485	2.810	2.520
	(1.886)	(1.716)	(1.736)	(1.729)	(1.784)	(1.723)
ln(Branches)	0.603**	0.612***	0.635***	0.633***	0.599**	0.588**
	(0.231)	(0.167)	(0.168)	(0.198)	(0.235)	(0.228)
Operating Expenses/Total Assets	60.96***	60.82***	61.56***	60.79***	61.91***	59.50***
	(12.67)	(11.69)	(11.40)	(12.27)	(12.08)	(12.18)
HHI	-21.01*	-23.31**	-22.34*	-7.500	-23.93**	-18.32*
	(11.35)	(9.907)	(12.29)	(12.94)	(10.59)	(11.03)
Observations	1312	1312	1312	1312	1312	1312
Within R-Squared	0.284	0.317	0.323	0.323	0.331	0.325
Between R-Squared	0.240	0.260	0.261	0.230	0.245	0.220
Overall R-Squared	0.196	0.214	0.213	0.186	0.208	0.182
Adjusted R-Squared	0.271	0.300	0.301	0.300	0.314	0.301

Table B.6: District wise MMC 3- Bank level

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Multimarket Contact Measure - 3	5.597***	4.053**	4.029**	4.555**	5.061***	4.540**
	(1.657)	(1.745)	(1.796)	(1.931)	(1.739)	(1.892)
NPA/Total Loans	-1.626	-1.035	-1.126	-1.213	0.263	-1.276
	(2.246)	(2.186)	(2.184)	(2.312)	(2.054)	(2.324)
Total Assets per Branch	0.00132	0.00492	0.00433	0.00294	0.00458	0.00309
	(0.00541)	(0.00759)	(0.00790)	(0.00746)	(0.00790)	(0.00719)
Total Loans/Total Assets	0.792	0.558	0.603	0.713	0.997	0.729
	(1.103)	(1.020)	(1.018)	(1.009)	(0.787)	(0.993)
Equity/Total Assets	2.719	2.646	2.524	2.345	2.655	2.377
	(1.872)	(1.706)	(1.723)	(1.717)	(1.769)	(1.712)
ln(Branches)	0.698***	0.735***	0.741***	0.790***	0.702***	0.750***
	(0.238)	(0.177)	(0.184)	(0.210)	(0.241)	(0.237)
Operating Expenses/Total Assets	60.90***	60.86***	61.63***	60.75***	62.16***	59.51***
	(12.64)	(11.63)	(11.33)	(12.18)	(12.11)	(12.11)
HHI	-20.86*	-21.90**	-21.57*	-8.484	-24.07**	-18.73*
	(11.60)	(10.42)	(11.70)	(12.89)	(10.80)	(10.83)
Observations	1312	1312	1312	1312	1312	1312
Within R-Squared	0.284	0.316	0.322	0.321	0.331	0.322
Between R-Squared	0.274	0.290	0.288	0.260	0.275	0.250
Overall R-Squared	0.228	0.243	0.239	0.215	0.238	0.210
Adjusted R-Squared	0.271	0.298	0.299	0.298	0.314	0.298

Appendix C

Appendix for Chapter 3

C.1 Data Description

I describe the definitions and sources of data for the various features used in the empirical estimation.

All household and individual features are defined using the survey data. Block 3 of the survey collects information on demographic characteristics at the individual level. This includes individual's relation to the household head, their gender and age, as well as their highest educational attainment. Crucially, information on holding deposit accounts in different sources, contributions to cooperative credit societies, SHGs/JLGs are also collected at the individual level in this block of the survey. Block 3 also collects information at the individual level on ownership of land (including agricultural land) as well as whether they hold a debit or credit card. Block 4 of the survey collects information at the level of the household, rather than the individual. It collects information on household size, religion, caste (social group)[1], land possessed as well as owned, and information on monthly expenditure for the household on various items including spending on durable goods in past year. Block 3 and 4 together allows one to construct individual and household specific demographic characteristics. Block 5 is broken down into two parts - the first part collects information on every plot of rural land owned, either exclusively or jointly, by the household as of 30-June-2018. It also captures data on whether female members of the household have ownership in any of the

[1]Castes or Social groups in India can be categorized into - Scheduled Castes, Scheduled Tribes, Other Backward Classes, Others (general including upper castes).

plots and their share in ownership[2]. The second part of Block5 also collects the same set of information but for every plot of urban land owned by a household. These plots of land could include irrigated/un-irrigated crop area, area for other agricultural or farm businesses, area for non-farm businesses, and residential areas including homestead. Block 6 collects information all buildings and constructions owned by the household as of 30-June-2018. This includes residential buildings (irrespective of whether the household lives in that building), buildings used for farm and non-farm business and other constructions (such as wells, bore wells, tube wells, field distribution system etc). We use information from blocks 5 and 6 to evaluate the value of immovable assets owned by each household. Households are then classified into quartiles based on the value of their immovable assets. The quartiles are defined within each state to account for any heterogeneity in valuations across states. Blocks 7-10 capture other assets owned by the household such as livestock and poultry, transport equipment, agricultural machinery and implements, and non-farm business equipment. We do not use data from this block in this analysis.

Blocks 11 and 12 are the main blocks of interest in this analysis. Block 11 collects information on all financial assets owned by the household as of 30-June-2018. The survey captures the value as of the survey data as well as the value as of 30-June-2018. The survey collects information on the source and form in which these assets are held. Block 12, on the other hand, captures information on cash loans payable by the household to institutional and non-institutional agencies as of the date of survey and transactions related to them between 01-July-2018 to the date of the survey. Information is collected on the interest rate, source of credit, tenure and purpose of the loan, as well as the outstanding amount as of 30-June-2018. Block 13 complements Block 12 by collecting information on any in-kind loan received by the household and its imputed monetary value.

Table C.2 presents the list of all features used related to individuals and household, and their

[2]We could not utilize the data on plot-wise ownership of land to due to sparsity of data.

data sources while Table C.1 lists the different state level features used in the analysis.

Table C.1

Feature	Level of granularity	Source
Number of branches managed by women	State	RBI - Handbook of Statistics on Indian States
Sex Ratio	State	RBI - Handbook of Statistics on Indian States
Rural Population (in 1000s, 2011)	State	RBI - Handbook of Statistics on Indian States
Urban Population (in 1000s, 2011)	State	RBI - Handbook of Statistics on Indian States
Net State Domestic Product (2018)	State	RBI - Handbook of Statistics on Indian States
Literacy Rate (2011)	State	RBI - Handbook of Statistics on Indian States
Per Capita Power (2018)	State	RBI - Handbook of Statistics on Indian States
Railway Route (kilometers, 2018)	State	RBI - Handbook of Statistics on Indian States
Roads (kilometers, 2018)	State	RBI - Handbook of Statistics on Indian States
Number of bank branches (2018)	State	RBI - Handbook of Statistics on Indian States
Amount of credit disbursed by commerical banks (INR million, 2018)	State	RBI - Handbook of Statistics on Indian States
Amount of deposits held in commerical banks (INR million, 2018)	State	RBI - Handbook of Statistics on Indian States
Crime rate against women (per 1000 women, 2017)	State	NCRB report, 2019
Enrollment rate - by gender and stage of school education (2018)	State	Ministry of Human Resource Development, Government of India

C.2 Robustness

In this appendix, we include results from alternative specification of empirical models for assessing household's access to finance. Inclusion of state fixed effects results in biased estimates in a Probit model. Therefore, we run a Logit model instead with state fixed effects given the dependent variable is a binary variable. The results from a Logit specification of the model are consistent with the results reported in Table 3.7.

Table C.2

Feature	Level of granularity	Source	Definition
Individual's age	Individual	AIDIS - Block 3	Response to Q4
Individual has deposit account (binary)	Individual	AIDIS - Block 3	Response to Q7-9
Individual holds debit or credit card (binary)	Individual	AIDIS - Block 3	Response to Q14
Individual Owns Land (binary)	Individual	AIDIS - Block 3	Response to Q11
Household has deposit account (binary)	Household	Derived	Based on whether atleast one individual in household has a deposit account
Household has debit or credit card (binary)	Household	Derived	Based on whether atleast one individual in household has a debit/credit card
Household owns land (binary)	Household	Derived	Based on whether atleast one individual in household owns land
Number of member in household with deposit account	Household	Derived	Based on number of individuals in household reporting to have a deposit account
Number of members of household with debit and credit card	Household	Derived	Based on number of individuals in household reporting to have a debit/credit card
Number of household members owning land	Household	Derived	Based on number of individuals in household reporting to own land
Age of household head	Household	Derived	Based on individuals response to Q3 and Q5 in Block 3
Size of the household	Household	AIDIS Block 4	Response to Q1
Number of children in household	Household	Derived	Based on responses to Q5 in Block 3
Dummy = 1 if highest education level in household is illiterate or below primary	Household	Derived	Based on responses to Q6 in Block 3
Dummy = 1 if highest education level in household is primary and above, upto higher secondary	Household	Derived	Based on responses to Q6 in Block 3
Dummy = 1 if highest education level in household is graduate and above	Household	Derived	Based on responses to Q6 in Block 3
Dummy = 1 if household's religion is Islam	Household	AIDIS Block 4	Response to Q2
Dummy = 1 if household's religion is Christian	Household	AIDIS Block 4	Response to Q2
Dummy = 1 if household's religion is Others (not Hindu)	Household	AIDIS Block 4	Response to Q2
Dummy =1 if household's caste is Scheduled Caste	Household	AIDIS Block 4	Response to Q3
Dummy =1 if household's caste is Scheduled Tribe	Household	AIDIS Block 4	Response to Q3
Dummy =1 if household's caste is Other Backward Caste	Household	AIDIS Block 4	Response to Q3
Dummy =1 if household's caste is Others	Household	AIDIS Block 4	Response to Q3
Dummy =1 if household's employment is Regular Salaried	Household	AIDIS Block 4	Response to Q4
Dummy = 1 if household's employment is Casual Labor	Household	AIDIS Block 4	Response to Q4
Dummy = 1 if household's employment is Self-employed	Household	AIDIS Block 4	Response to Q4
Dummy = 1 if household's employment is Others	Household	AIDIS Block 4	Response to Q4
Dummy = 1 if household's residence is Urban	Household	AIDIS Block 1	
Household's usual monthly total expenditure (INR)	Household	AIDIS - Block 4	Response to Q10.5
Value of land owned by households in rural area (INR)	Household	Derived	Derived from responses in Block 5.1
Value of land owned by households in urbanl area (INR)	Household	Derived	Derived from responses in Block 5.2
Value of buildings owned by household (INR)	Household	Derived	Derived from responses in Block 6
Value of households immovable assets (INR)	Household	Derived	Derived from responses in Block 5.1, 5.2, and 6
Total amount of cash held by household (INR)	Household	AIDIS Block 11A	Response in row 1
Total amount of deposits held in Banks by household (INR)	Household	AIDIS Block 11A	Response in rows 3 and 4
Total amount of deposits held in Non-Bank formal sources by household (INR)	Household	AIDIS Block 11A	Response in rows 5-9
Household has atleast one loan from formal sources (binary)	Household	Derived	Derived from responses in Block 12
Household has atleast one loan from bank (binary)	Household	Derived	Derived from responses in Block 12
Household has atleast one loan from non-bank formal sources (binary)	Household	Derived	Derived from responses in Block 12
Household has atleast one loan from informal sources (binary)	Household	Derived	Derived from responses in Block 12
Total amount of formal loans availed by household (INR)	Household	Derived	Derived from responses in Block 12
Total amount of informal loans availed by household (INR)	Household	Derived	Derived from responses in Block 12
Household has atleast one loan in-kind (non-cash) (binary)	Household	Derived	Derived from responses in Block 13
Total value of in-kind loan availed by household (INR)	Household	Derived	Derived from responses in Block 13

Table C.3

Outcome of interest	Household has access to formal credit	Household has access to formal credit - Banks	Household has access to formal credit - non bank entities	Household has access to informal credit	Household has access to in-kind loans
Dummy =1 if hh head is female (0 otherwise)	-0.16***	-0.29***	0.12***	-0.05**	0.08**
	(0.021)	(0.025)	(0.028)	(0.022)	(0.039)
Age of hh head	0.05***	0.04***	0.05***	0.04***	0.02***
	(0.003)	(0.004)	(0.005)	(0.003)	(0.006)
Size of household	0.09***	0.03***	0.10***	0.07***	0.01
	(0.006)	(0.006)	(0.008)	(0.006)	(0.011)
Dummy =1 if max educ in hh is between primary and higher secondary	0.18***	0.16***	0.18***	0.08***	0.12***
	(0.025)	(0.030)	(0.033)	(0.024)	(0.043)
Dummy =1 if max educ in hh is grad and above	0.24***	0.37***	-0.10**	-0.28***	-0.05
	(0.029)	(0.034)	(0.041)	(0.030)	(0.054)
Dummy =1 if hh religion is Islam	-0.18***	-0.17***	-0.11***	0.10***	0.28***
	(0.023)	(0.025)	(0.031)	(0.023)	(0.038)
Dummy =1 if hh religion is Christian	-0.16***	-0.15***	-0.02	-0.02	0.04
	(0.039)	(0.044)	(0.059)	(0.040)	(0.098)
Dummy =1 if hh religion is Others (Sikhism, Jainism, Buddhism, Zoorastrianism)	0.06	0.07	0.04	-0.06	0.30***
	(0.040)	(0.043)	(0.060)	(0.043)	(0.072)
Dummy = 1 if hh caste is SC	0.06***	-0.15***	0.30***	0.19***	0.20***
	(0.022)	(0.025)	(0.029)	(0.023)	(0.040)
Dummy = 1 if hh caste is ST	-0.10***	-0.31***	0.21***	-0.10***	-0.30***
	(0.026)	(0.030)	(0.034)	(0.028)	(0.053)
Dummy = 1 if hh caste is OBC	-0.04**	-0.09***	0.10***	0.19***	0.19***
	(0.017)	(0.019)	(0.024)	(0.018)	(0.033)
Dummy = 1 if hh identifies as casual employed	-0.16***	-0.41***	0.18***	0.30***	0.16***
	(0.021)	(0.025)	(0.029)	(0.023)	(0.044)
Dummy = 1 if hh identifies as self employed	0.18***	0.14***	0.13***	0.31***	0.23***
	(0.018)	(0.020)	(0.026)	(0.021)	(0.038)
Dummy = 1 if hh identifies as Others for employment	-0.47***	-0.37***	-0.45***	-0.23***	-0.21***
	(0.030)	(0.034)	(0.047)	(0.034)	(0.063)
Dummy = 1 if hh is in Urban area	0.08***	0.07***	0.04*	-0.35***	-0.33***
	(0.015)	(0.017)	(0.022)	(0.016)	(0.031)
Second asset quartile	0.22***	0.35***	0.01	-0.01	-0.02
	(0.020)	(0.025)	(0.026)	(0.020)	(0.036)
Third asset quartile	0.46***	0.70***	-0.05*	-0.05**	0.02
	(0.020)	(0.025)	(0.028)	(0.021)	(0.038)
Fourth asset quartile	0.68***	0.99***	-0.17***	-0.19***	-0.01
	(0.022)	(0.026)	(0.031)	(0.023)	(0.042)
Second quartile - monthly expenditure	0.25***	0.24***	0.15***	0.13***	0.03
	(0.020)	(0.024)	(0.028)	(0.021)	(0.038)
Third quartile - monthly expenditure	0.33***	0.38***	0.12***	0.09***	0.00
	(0.021)	(0.025)	(0.030)	(0.023)	(0.041)
Fourth quartile - monthly expenditure	0.47***	0.60***	0.01	0.01	0.04
	(0.024)	(0.027)	(0.035)	(0.026)	(0.047)
Dummy =1 if hh has deposits with bank	0.90***	0.96***	0.56***	0.16***	0.06
	(0.045)	(0.062)	(0.059)	(0.041)	(0.074)
Dummy =1 if hh has deposits with non-bank formal sources	0.14***	0.10***	0.10***	0.18***	0.08*
	(0.024)	(0.026)	(0.033)	(0.026)	(0.047)
State FE	Yes	Yes	Yes	Yes	Yes
Observations	111,508	111,508	111,508	111,508	111,175
Pseudo R square	0.0802	0.102	0.0681	0.0788	0.0640

Note: Robust standard errors in parentheses (clustered at household level); *** p-value < 0.001, ** p-value < 0.05, * p-value < 0.001