



RESEARCH ARTICLE

Modelling credit and savings behaviour of chit fund participants [version 1; peer review: 1 approved, 1 approved with reservations]

Preethi Rao , Sharon Buteau

IFMR LEAD, Institute for Financial Management and Research, Nungambakkam, Chennai, 600006, India

V1 First published: 04 May 2018, 2:26
<https://doi.org/10.12688/gatesopenres.12767.1>
Latest published: 04 May 2018, 2:26
<https://doi.org/10.12688/gatesopenres.12767.1>

Abstract

Chit funds are a form of Rotating Saving and Credit Association (ROSCA) prevalent in India. In this institution, a group of individuals pool in equal amounts of money at a fixed frequency, and at every time period a round of competitive bidding takes place among the individuals to identify a borrower for the collected amount. The borrower foregoes his/her right to participate in further auctions, but continues paying his/her share of the pool till every individual in the group has collected the pooled amount. The auction process relies on the bidders' willingness to give up a certain amount of the pool and take the rest as loan. The amount given up at every time period is shared equally among all individuals in the group.

A chit fund, therefore, is structurally similar to the modern formal banking and financial institutions: they act as an intermediary to optimally mobilize funds collected from savers to borrowers and manage repayment of loans from borrowers such that savers receive their dues at the appropriate time. The latter role exposes the industry to credit risk.



Modern banking and financial institutions rely heavily on their ability to assess and mitigate credit risk; and, over the last century, they have been able to move from a system where risk assessment was based on human judgment and simple intuition to a system reliant on statistical and mathematical techniques. However, the standard chit fund company still relies heavily on human judgment for assessing risk. The pilot envisages a move towards a standardized statistical methodology for risk assessment and mitigation as the path ahead for the industry.

Keywords


Credit, Prediction, Data

Open Peer Review

Approval Status  

	1	2
version 1		
04 May 2018	view	view

1. **Yu Zhang**, Eastern Michigan University,
Ypsilanti, USA

2. **Stefan Klonner** , Heidelberg University,
Heidelberg, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Preethi Rao (preethi@ifmr.ac.in)

Author roles: **Rao P:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Buteau S:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Bill and Melinda Gates Foundation [OPPGD489].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Rao P and Buteau S. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Rao P and Buteau S. **Modelling credit and savings behaviour of chit fund participants [version 1; peer review: 1 approved, 1 approved with reservations]** Gates Open Research 2018, 2:26 <https://doi.org/10.12688/gatesopenres.12767.1>

First published: 04 May 2018, 2:26 <https://doi.org/10.12688/gatesopenres.12767.1>

Table of Contents

1. Background

- 1.1 Why credit scoring?
- 1.2 Application of credit scoring to the chit fund industry

2. Research design and methodology

- 2.1 Analysis outline
- 2.2 Data collection process
- 2.3 Partnering companies and data collected from them

3. Descriptive statistics

- 3.1 Overall summary
- 3.2 Demographics
- 3.3 Member participation
- 3.4 Irregular repayment behaviour
- 3.5 Collateral and surety information
- 3.6 Auction information

4. Credit scoring model

- 4.1 Where does risk need to be assessed?
- 4.2 Risk assessment at the point of entry
- 4.3 Risk assessment at the point of claiming prized amount as loan

5. Regression results

- 5.1 Regression analyses at the time of joining
- 5.2 Derivation of credit score at the time of joining
- 5.3 Regression analysis at the time of taking loan
- 5.4 Derivation of credit score at the time of borrowing

6. Application of findings to the industry

7. Scalability

8. Conclusions

List of Tables

Table 1: Overview of types of schemes received

Table 2: Trends Observed in Chit Industry

Table 3: Member Demographics

Table 4: Types of Occupation

Table 5: Member Participation

Table 6: Details of Repeat Members

Table 7: Repayment Behaviour

Table 8: Duration of Defaults (in months)

Table 9: Default Rates by Chit Value

Table 10: Default Rates by Chit Duration

Table 11: Default Rates by Monthly Contribution

Table 12: Default Rates by Monthly Contribution-Chit Duration Combination

Table 13: Seasonality in Late Payments and Defaults

Table 14: Collateral and Surety Information

Table 15: Participation in Auctions

Table 16: Unmet demand for liquidity in 30 month schemes

Table 17: FICO Model and Application to Chit Fund Industry

Table 18: Occupation and Income Level Categories and Variables

Table 19: Regression Part I: OLS regression at the Time a Client Joins the Company

Table 20: Scores Derived for Overall Riskiness

Table 21: Scores Derived for Riskiness after Winning Prized Amount

Table 22: Final Scores at the Time of Joining

Table 23: Regression Part II: Probit Regression at the Time of Borrowing

Table 24: Scores Derived for Riskiness at Time of Borrowing

Table 25: Final Score at the Time of Borrowing

1. Introduction

1.1 Why credit scoring?

Banks and other financial institutions face myriad types of risk. The primary form of risk attributed to a banking or financial institution is credit risk—the risk associated with a borrower becoming unable to repay his/her debts to the institution. Therefore, a crucial pillar in the existence of the modern banking and financial industry is its ability to ascertain the risk attributed to it by its borrowers, and making decisions to mitigate the same.

At the onset of the modern banking and financial institutions, human judgement was the prevalent form of assessing the potential risk attributable to a creditor¹. Decisions were based on previous clientele of the lenders and a simple intuition-based system. Credit denied or granted in such a manner proved arbitrary and was solely based on the ‘gut-feel’ of the agent of the institution, who faced the prospective creditors. Over the last century, a systemic change in this process has been observed with the industry moving towards a steady reliance on statistical and mathematical techniques to assess credit risk.

This move has slowly eradicated the arbitrariness of human judgement and ushered in a system of methodological rating of credit through credit bureaus and credit scoring, whereby a prospective borrower is judged on the basis of his/her individual financial and socio-economic characteristics, rather than what the lender perceives to be historically similar behaviour to prior risky borrowers².

However, in developing countries, the absence of efficient credit bureaus makes it costly for financial institutions to collect information on clients’ credit-worthiness. Furthermore, because they rely on the relationship with their customers to assess their risk, they often do not feel the necessity.

The chit fund industry is one such indigenous financial system that mostly relies on personal judgement assessments and contacts to assess the potential clients. Chit funds are a form of Rotating Saving and Credit Association (ROSCA) prevalent in India. In this institution, a group of individuals pool in equal amounts of money at a fixed frequency and at every time period a round of competitive bidding takes place among the individuals to identify a borrower for the collected amount. The borrower foregoes his/her right to participate in further auctions, but continues paying his/her share of the pool till every individual in the group has collected the pooled amount. The auction process relies on the bidders’ willingness to give up a certain amount of the pool and take the rest as loan.

The amount given up at every time period is shared equally among all individuals in the group. In the event that multiple members bid at the cap, a lottery is conducted amongst the highest bidders to decide the winner³.

In comparing the chit fund industry with the modern banking and financial institutions, many similarities can be drawn up. In their functions, like the banking industry, chit funds act as an intermediary to optimally mobilize funds from savers and channelize them to borrowers, and manage the repayment of loans from the borrowers so that savers may receive their dues when they make a claim⁴. Our initial field studies show that the chit fund industry too faces credit risk as their primary source of risk, much like the formal banking and financial institutions. However, over the many years that the chit fund industry has existed in the market, it has been unable to move to a standardized statistical method of assessing risk and still relies primarily on human judgement.

As observed in the banking industry, a move towards standardized statistical methods in ascertaining risk accruable to clients may prove to be crucial in establishing the chit fund industry as a credible formal financial institution and enable it to grow in a more systematic and balanced manner⁵.

1.2 Application of credit scoring to the chit fund industry

In the setting up of a typical chit fund, two primary tasks exist—one, to evaluate the credit worthiness of prospective clients and, two, assess the borrowing and saving requirements of the prospective clients so that liquidity is provided optimally throughout the duration of the scheme.

The role of any statistical process in the chit industry, therefore, would be two-fold: first, to assess the riskiness of a prospective client at the time of entry and, second, to assess whether a person is likely to be a borrower or saver and match him/her to an appropriate group. Thus, more than just a credit scoring process, the problem faced by the chit fund is an optimization problem—minimize the total credit risk of the group, and match prospective savers and borrowers such that demand and supply of liquidity is balanced.

In addressing this problem, good information about client characteristics is a necessity. Currently, chit funds function primarily on social networks⁶. Further, our primary field studies indicate that each client is introduced by existing clients. All the decisions made by the company agents, with regards to assessing

¹R.B. Avert, P.S. Calem and G.B. Caner, ‘An Overview of Consumer Data and Credit Reporting’, *Federal Reserve Bulletin*, February 2003.

²E. Huang and C. Scott, *Credit Risk Scorecard Design, Validation and User Acceptance*, Credit Scoring Conference – The Credit Research Centre, University of Edinburgh, 2007.

³The Chit Fund Act, 1982, stipulates that the bid amount should not exceed 40 per cent of the chit value.

⁴P. Rao, ‘Chit Funds – A Boon to Small Enterprises’, Small Enterprise Finance Centre, Institute for Financial Management and Research Working Paper Series, January 2007.

⁵A. Estrella *et al.*, ‘Credit Ratings and Complementary Sources of Credit Quality Information’, Bank for International Settlements: Basel Committee for Banking Supervision Working Papers No. 3, August 2000.

⁶J. Eeckhout and K. Munshi, ‘Matching in Informal Financial Institutions’, *Journal of European Economic Association*, September 2010.

credit worthiness or the borrowing or saving nature of the client rely on the relationship the introducer has with the company. Surprisingly, most chit fund companies do not collect in-depth socio-economic information of such clients, and are satisfied with the private information collected over the introduction.

From a financial inclusion point of view, these social networks appear to have great value—they inherently reduce the stringency of requirements for borrowers and savers when compared to the formal banking and financial institutions, thereby allowing access to financial products for those who were deemed unfit by the formal institutions⁷. Initial field studies show that this unique feature, however, has not been maximized upon by the chit fund industry due to the heavy reliance on human judgement for decision-making. The lack of a standardized risk assessment creates a residual risk in the selection of groups, since officers might vary in their ability to assess the credit and savings behaviour of clients, and a departure of an officer from the company would result in the loss of valuable information for the business.

The need of the hour, therefore, is a systematic collection of financial and socio-economic information from all prospective clients, and a standardized assessment and evaluation of said information to rate the credit worthiness of a client. This process would create an automated and uniform system of assessing prospective clients, supplementing the judgement of company officers, and reducing the risk exposure of the company. A reduction in the risk exposure would result in an increase in the acceptance rate of potential clients and, consequently the customer base of the company.

In order to assess the creditworthiness of clients and their savings and borrowing needs, chit fund companies should ideally shift their current primary reliance on relationship-based client assessment to a system where data-based models supplement existing assessment methods for reasons enumerated earlier. Hence, in this project, we analyse how client information, their retrospective transaction data, and their track records in the bidding process can be optimally used to predict their creditworthiness and financial needs. In doing so, the project aims to be helpful in advising and guiding chit fund companies on how to systematically collect, use, and evaluate client data to efficiently assess a participant's credit worthiness.

In summary, the main objectives of this pilot were as follows:

1. Assessing the existing client information, bidding and repayment data collected by the chit fund companies.
2. Producing a prediction model (for old and new clients) that will create a credit score and describe how personal characteristics can be used to assess the creditworthiness, borrowing, and savings needs of clients.

This allows devising of simple weighting algorithms based on the prediction models that chit funds can choose to use in designing the composition of a chit scheme.

The following sections elaborate on the data collection process, description of the data, the credit scoring model and the potential of scalability of the model, as well as the role of technology in reaching these objectives.

2. Research design and methodology

2.1. Analysis outline

As stated in the previous section, this pilot aimed to design an algorithm that would analyse the demographic and socio-economic information, as well as the financial track record of old and prospective clients to predict their credit worthiness and savings or borrowing needs in order to optimize group formations.

In order to achieve the above stated goal, three primary tasks were carried out. First, there was an assessment of client information collected by companies; this included demographic and other socio-economic information, transaction records, and bidding information.

Second, the available information was used to design a predictive model that would be instrumental in measuring the risk accruable to the chit fund company and the nature of the client's repayment characteristics.

And finally, results obtained from above were used for evaluating how the quality of predictions can be improved with improved data collection methods.

The overall analysis is retrospective, that is, all analyses were carried out using historical data. The data was collected from a group of five different companies located in various parts of Tamil Nadu, Andhra Pradesh, and Delhi. The time range for which the data was available ranged from 5 to 13 years across different companies. After sifting the complete data for the best possible collection of information, the payment records of about 5,000 individuals were tracked over a period of 13 years from two different companies.

2.2. Data collection process

All companies maintain transactional administrative data in two forms: primary transaction data, stored in customized database programs, and secondary client based information, usually stored in physical form.

The administrative data is collected in a four-step procedure by the company:

1. The first step takes place when a new member joins the company. At this point the member is required to fill an application form, from where basic socio-economic and demographic information, like age, sex, occupation, address, and monthly income are collected. This is recorded electronically.

⁷S. Bateau and P. Rao, *Chit Funds as an Innovative Access to Finance for Low-income Households*, Small Enterprise Finance Centre, Institute for Financial Management and Research, August 2010.

2. The second step occurs once a month, following the monthly collection of subscription amounts and auctioning of the pot. At this point, the company records its transactions with the member. In this step, transactional information like subscription amount paid, dividend earned, fines paid, prized amount won, and details of mode of payment are recorded by the company. This too is recorded electronically.
3. The third step occurs during the auctioning of the pot, when the company records the list of members present for auctioning, their bid amounts, the members' modes of bidding—whether they were present at the auction or whether they bid by letter or by agent, the winning bid amount, and whether the winner was selected by lottery. This information is collected manually.
4. The final step occurs when the member wins the auction and is eligible for the prized amount. Before collecting the prized amount the member is expected to furnish a list of guarantors, or collaterals that total up to the amount the member is liable for, and basic identity proofs that satisfy pre-existing norms set down by the company. If the liability of the member is greater than a relatively large benchmark amount, the member is required to provide both guarantors and collaterals. The furnished information is recorded in a manual form by the company, but in most cases, a summary of the details are recorded electronically.

The above mentioned steps can be understood better with the graphical illustration presented in [Figure 1](#). In this example the scheme duration is 20 months, and the member wins the auction in the sixth month.

Based on our field studies, we infer that most companies in the chit fund industry in India are, by nature, family-owned and cater to an exclusive set of clients. Their clients belong to closed networks, with a vast majority of new clients being introduced by existing clients of the companies. This results in a

personal relationship and mutual trust between the client and the company.

Some companies, however, have deviated from the family run business model and transformed the industry with their large-scale corporate model for chit funds, and have been conducting business based on this philosophy over the last decade. This model relies heavily on agents who form an added layer of bureaucracy and act as intermediates between the client and the company.

Both models have their unique shortcomings as sources of data. For example, while the information collected through guarantor and collateral is of high value in the second model, the formality of the corporate structure forces the companies to return all forms of documentation, which could be an important source of demographic information for this study, back to the clients upon successful termination of a chit scheme. This makes a large quantity of personal demographic information unavailable. Similarly, in the family-run model, there is flexibility in the number and type of personal documents collected at the time of joining and the number of guarantors, lassitude in the recording of late payments, and management of ledger information which display irregularities that needed large-scale manual editing.

Therefore, the data collected for the study had to be extracted from the documentation created at the four steps outlined above. All physical records were manually entered by conducting data entry operations at the location of the records. Several tables thus obtained were collated, cleaned, and combined to form the final dataset. This dataset is in a panel format, where all information is organized by name of chit schemes and by name of members of each scheme.

2.3. Partnering companies and data collected from them

In our research design, we planned the analysis to be carried out on a dataset that is representative of the broad spectrum of formal chit fund users across the country. In line with this mandate,



Figure 1. An example of the data collection process at a typical chit fund company, where the member participates in a 20-month scheme and wins the auction in the sixth month.

the initial intention was to find two companies that maintain good quality database and represent a large section of the chit fund industry in terms of the spectrum of members.

Therefore, companies in different states across the country—Tamil Nadu, Delhi, Andhra Pradesh, and Maharashtra were approached. Following some initial resistance, a large number of companies were willing to share their data with us. Owing to this response, we collected data from five different companies from the above mentioned states from which, after further scrutiny, focused our analyses of data from two of the companies⁸. The data collected from the remaining companies will be used to assess scalability and to understand the future requirements from the chit fund companies in order to implement a practical credit scoring model. A point of learning for us, and perhaps the chit fund industry as a whole, from this extended data collection process is that there exists an urgent need for companies to have a standardized process for data collection and management such that they will have access to greater information about their members and, consequently, have a higher level of adaptability when exposed to random risks from the members.

3. Descriptive statistics

3.1. Overall summary

As stated in the previous section, the final dataset was extracted from two companies. This data covers an operational period of 15 years from 1996 to 2011. It consists of the transaction records of almost 5,000 individuals who have participated in a total of 274 chit schemes, each ranging from duration of 20 months to 50 months. The final format of the dataset is set in a panel form, thereby allowing the observer to follow all transactions made by each individual over the duration of the schemes that they were participating in. Over the next few pages, with the aid of descriptive statistic tables, we will draw an outline of the data collected from the companies and the basis formed by these statistics for further in-depth multivariate regression analyses.

Table 1 outlines the distribution of the schemes which constitute the dataset. This gives an overview of the characteristics of the dataset being used for the analysis.

From **Table 2**, it can be seen that over the last 15 years: (1) the proportion of high value schemes has been increasing; (2) the proportion of schemes with high monthly contribution has been rising; and (3) the proportion of shorter duration schemes has been rising. This implies that the focus segment of population for the companies has been transforming over the years, the chit industry is moving away from lower and middle income groups, and seeking to primarily target higher income groups. From interviews in the field, the principal reason for the above trends is that the transaction costs for running schemes of all values remains the same, thus higher

value schemes allow the companies to make greater returns on their investments.

As discussed in the earlier sections, the chit fund model is highly applicable in the context of targeting financial inclusion for lower income households due to their social/relationship network lending and saving characteristics. Unfortunately, it is evident from the above mentioned trends, in order to target the financial inclusion goal, a certain amount of intervention is needed.

Both companies maintained a closed network of members. Though they maintained, what at first appears to be a near complete demographic record of their clients, a lack of regular updating of the databases restricted the information of clients available to that of what was available at their time of joining. Given the fact that a vast majority of clients stay on with the companies for extended lengths of time, the demographic information aspect of the dataset needs to be improved upon in the future.

3.2. Demographics

The following tables give us an indication of the characteristics of the demographic and socio-economic parameters of the members, in terms of gender, age, and stated income levels, as collected by the companies at the time of the members' joining.

As observable in **Table 3**, the gender ratio of members of the companies maintains an overall approximate average of 60 males for every 40 females. This figure might not be truly indicative of the level of female participation in the companies, since, as stated by the company representatives, in the case of many female participants the tickets were dummy tickets—bought by a male member of the family in the name of female members. This usually occurs when the male member of the family is the only earning member and saves in the name of other members of the family.

The average age of the representative population also remains steadily above 40 years. This is indicative of how the target population of the chit industry continues to be middle-aged individuals. Combining this trend with the trends observed in **Table 2**, the industry seems to be targeting middle-aged individuals who fall in the higher-income groups. As per the trends observed in the Indian population growth, a majority of the population lies in the age group 15–30; therefore, future targeting of younger populations might prove to be the much needed shot-in-the-arm for the chit fund industry⁹.

The reported income levels of the members appear to be grossly understated. As per **Table 2**, if a large proportion of schemes have monthly contributions of greater than Rs.3,000, then, assuming contributions towards chit funds to be a form of savings and an average national savings rate of around 30 per cent, it is highly unlikely that average income levels

⁸The choice of companies was based on the quality of data maintained by the companies. Summary of choice of companies is shown in tables A1 and A2 in [Supplementary File 1](#).

⁹Planning Commission of India, *Population Growth – Trends, Projections, Challenges and Opportunities*, Planning Commission of India, 2000.

Table 1. Overview of types of schemes received.

Total Number of Companies	2
Total Number of Schemes	274
Total Number of Members	4,936
Total Number of Tickets	9,318
Average Number of Tickets per Member	1.9
Time Period Covered in Dataset	1996 to 2011
Range of Chit Values	Rs.15,000 to Rs.1,000,000
Range of Monthly Contributions	Rs.500 to Rs.25,000
Range of Scheme Duration	20 months to 50 months

Note: A detailed summary of the distribution of chit schemes by year of starting is available in [Supplementary File 1¹⁰](#).

Table 2. Trends Observed in Chit Industry.

Starting Year	Proportion of Schemes above Rs.1,00,000 (%)	Proportion of Schemes with Monthly Contribution above Rs.3,000 (%)	Proportion of Schemes shorter than 30 months (%)
1996	11.8	5.9	41.2
1997	23.1	7.7	38.5
1998	33.3	11.1	33.3
1999	21.1	21.1	57.9
2000	36.4	27.3	40.9
2001	41.4	34.5	62.1
2002	43.5	85.5	65.2
2003	59.1	40.9	50.0
2004	68.0	48.0	56.0
2005	64.3	50.0	60.7
2006	56.0	52.0	84.0
2007	66.7	46.7	80.0
2008	58.3	41.7	91.7
2009	66.7	66.7	(100.0)
Average	46.4	38.5	58.6 [*]

Note:* For the last average figure, that is, average proportion of schemes less than 30 months in duration, the figure for 2009 has not been included.

can remain around Rs.12,000. For the future, it might be useful to collect correct income levels, or update this field regularly, since the disposable income of an individual may determine whether an individual can make the monthly contributions.

The observable trend of greater participation in higher value schemes, and grossly understated income levels, might imply

that users are channelling unaccounted earnings into the chit fund industry.

Table 4 shows the type of occupations reported by the members. Private sector employees figure as the largest group, forming

¹⁰Complete table available in Table A3 in [Supplementary File 1](#).

Table 3. Member Demographics.

Year of Joining	Gender			Age (in years)			Stated Income (Rs)		
	%age of data available	Male	Female	%age of data available	Mean	Standard Deviation	%age of data available	Mean	Standard Deviation
1996	63.1	65.8	34.2	90.6	40.7	12.4	81.8	6,893.1	8,558.1
1997	99.8	61.4	38.6	95.8	42.0	12.7	82.9	8,515.7	7,438.1
1998	99.8	63.7	36.3	93.1	42.1	12.9	80.4	9,670.6	9,856.8
1999	100.0	66.9	33.1	91.1	41.5	12.7	71.0	8,729.6	6,247.4
2000	100.0	65.3	34.7	92.3	41.5	12.4	73.0	10,823.4	8,498.0
2001	100.0	60.6	39.4	91.6	41.3	12.1	69.4	11,851.2	9,682.1
2002	100.0	64.5	35.5	91.9	41.3	13.0	71.7	12,195.4	8,176.7
2003	100.0	59.1	40.9	91.3	40.7	12.6	69.1	12,872.2	12,311.9
2004	100.0	58.9	41.1	87.9	41.3	12.4	60.7	14,387.7	11,465.4
2005	99.8	64.3	35.7	86.8	40.9	11.9	64.0	12,314.5	8,625.8
2006	99.7	62.7	37.3	83.9	41.2	11.4	63.7	15,463.9	18,847.5
2007	100.0	62.6	37.4	88.2	42.4	11.7	77.9	14,424.8	15,257.1
2008	100.0	52.7	47.3	90.2	42.2	11.2	96.7	16,470.2	13,847.4
2009	100.0	78.9	21.1	95.2	38.8	8.9	95.2	17,700.0	13,727.0
Overall	95.5	57.6	42.4	90.6	41.3	12.6	72.7	12,308.0	10,350.4

Table 4. Types of Occupation.

Occupation Type	Number	Percentage (%)
Private Service	2,218	44.9
Business	770	15.6
Others/Unstated Occupation	649	13.1
Housewife	406	8.2
Small Business Owner	301	6.1
Government Service	298	6.0
Retired	214	4.3
Self-employed Professional	80	1.6

roughly 45 per cent of the chit fund users. Self-employed professionals constitute the smallest proportion of members at less than 2 per cent.

3.3. Member participation

Table 5 outlines the participation of members across the two selected companies. This is an interesting statistic to take note of, since the chit fund industry relies on social networks as a means of expanding business. Both companies show a significant number of members who have returned to participate in schemes after participating in one scheme.

Table 6 describes the nature of the repeat customers observed in Table 5. On average, members wait for approximately 20

months before joining a new scheme, and those customers who participate only in one scheme at a time, wait for nearly a year after completion of one scheme before joining the next. A particular point of interest is how a large proportion of repeat customers participate in multiple schemes at the same time: almost 70 per cent of the members, who participate in multiple schemes during their tenure at the companies, participate in more than one scheme at the same time.

3.4. Irregular repayment behaviour

Table 7 outlines the rates of irregularity in payments in the dataset. By definition of the chit industry, defaulters are those individuals who have not paid their dues for more than 3 months. Default behaviour can be observed both before and

Table 5. Member Participation.

Number of Schemes	Company 1		Company 2	
	Number of Participants	Percentage of Participants (%)	Number of Participants	Percentage of Participants (%)
1	3017	67.2	192	43.1
2	785	17.5	96	21.6
3	318	7.1	56	12.6
4	157	3.5	22	4.9
5+	214	4.8	79	17.8

Table 6. Details of Repeat Members.

Number of months between starting two schemes	
Average	19.6
Standard Deviation	16.5
Minimum	0.1
Maximum	119.7
Number of months between ending one scheme and starting the next	
Average	10.1
Standard Deviation	12.8
Minimum	0.1
Maximum	70.1
Participating in more than one scheme at the same time	
Number	1223
Percentage*	70.8%

* For calculating the percentage, the reference group is the total number of members across both companies who participate in more than one scheme.

Table 7. Repayment Behaviour.

Type of Behaviour	By Members (%)			By Total Transactions (%)		
	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning
Defaults	35.1	20.4	23.6	4.0	2.0	2.1
Late Payments	86.2	66.9	70.6	16.3	6.9	9.4
Early Payments	98.6	95.9	91.7	72.3	38.4	33.9
Part Payments	83.3	60.8	69.3	12.0	4.9	7.2
Lump-sum Payments	99.8	96.3	89.3	19.5	9.4	10.0

Note: Defaults by members are calculated as the percentage of members out of the total members, who have defaulted at least once during the time that they participated in schemes. Default by total transactions is calculated as the number of times a member defaults out of all the transactions made by him/her. The same rules apply to the other types of irregular payment behaviours that are considered.

after a member takes the loan. The industry focuses primarily on the group of individuals who default after taking the loan, but both type of defaulters cause the company to lose money, since the company is forced to pool in working capital to ensure smooth functioning of the schemes for the long duration of time that the member defaults for.

In this case, it is evident that more than 35 per cent of the members have defaulted at least once at any point of the scheme; but, of interest to us is the nearly 24 per cent of members who have defaulted after taking the loan. Though the overall rate of default transactions remains at 4 per cent, and 2 per cent for the period after taking a loan, being able to understand what drives this payment behaviour and tackle the same is of utmost concern in mitigating risks faced by the companies.

Other than the default behaviour, we also observe a large proportion of members who make irregular payments. Late payments are those payments which are made after the scheduled time, but less than 3 months from the due date. Early payments are those payments which are made before the due date. Part payments are those payments which are partial payment of dues and lump-sum payments are those which are bulk payments for dues of multiple months.

Late payments, part payments, and lump-sum payments also add to the credit risk faced by the company. Although they do not attribute as much risk as defaults after a loan, their extremely high incidence is a cause of worry. Almost 90 per cent of members have made at least one late payment, and close to 85 per cent of the members have made at least one part payment. Early and lump-sum payments are made by almost the entire population. Summarily, irregularity in receipts from the members is of high concern.

Table 7, in combination with Table 8, describes the severity of defaults when they occur.

This table demonstrates that when a default occurs, that is, when the member has been unable to pay dues for more than 90 days they are unable to clear dues for another 2 months on average. This obviously results in the company having to provide excess working capital to the schemes where such defaults occur, thereby increasing the cost of risk mitigating.

Table 8. Duration of Defaults (in months).

	Overall	Before Winning	After Winning
Average	2.4	1.8	2.8
Standard Deviation	2.7	2.1	3.1
Minimum	1	1	1
Maximum	37	23	37

In the following tables, we observe how default rates differ among different classes of chit funds. In this initial analysis, we consider four classes based on chit scheme value, chit scheme duration, monthly contribution level, and a combination of monthly contribution and chit scheme duration:

1. Class-based on chit scheme value: High value schemes are those schemes that are of value greater than Rs.1,10,000 and low value schemes are all schemes below that level.
2. Class-based on chit scheme duration: Long duration schemes are those schemes which are of length greater than 30 months and short duration schemes are all schemes shorter than that level.
3. Class-based on monthly contribution level: High value monthly contribution schemes are those schemes with monthly contributions greater than Rs.3,000 and low value monthly contribution schemes are those schemes with monthly contributions lesser than that value.
4. Class-based on combination of monthly contribution and chit scheme duration: Here we consider true high value, that is, schemes with high monthly contribution level and short chit scheme duration.

It appears that there is not much difference in the default rates of high and low value schemes, as evident from Table 9. In both cases, percentage of members who have defaulted at least once, overall, remains approximately 34 per cent and the percentage of total transactions that were default transactions are around 4 per cent (Table 10).

Longer duration schemes indicate higher rates of default, both in terms of percentage of members (around 38 per cent vs. 28 per cent) and number of transactions (around 5 per cent vs. 3 per cent). This is an expected result since longer schemes will allow more people to default. Therefore, assuming long duration schemes to be of higher risk based on this statistic alone might be inconclusive.

Table 11 exhibits similar trends as the defaults in the chit value case. It appears that there is not much difference between high monthly contribution and low monthly contribution schemes in terms of percentage of members defaulting. In both cases, roughly 30–35 per cent of the members default at least once. However, it seems that the rate of default transactions in high value monthly contribution schemes are more than one percentage point higher than the low value monthly contribution schemes. Therefore, monthly contribution levels might have a greater role in defining default categories.

An interesting class of schemes run by the companies are true high value schemes. These are schemes that have monthly contribution greater than Rs.3,000 and are lesser than 30 months in duration. If we study Table 12, it appears that true high value schemes differ significantly from the overall high

Table 9. Default Rates by Chit Value.

By Member (%)						By Total Transactions (%)					
High Value			Low Value			High Value			Low Value		
Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning
33.7	20.1	22.1	33.9	19.6	22.7	3.6	1.8	1.8	4.1	2.0	2.1

Note: High value schemes are those schemes that are of value greater than Rs.1.1 lakh, and low value schemes are those which are below this benchmark.

Table 10. Default Rates by Chit Duration.

By Member (%)						By Total Transactions (%)					
Long Duration			Short Duration			Long Duration			Short Duration		
Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning
37.9	22.6	25.0	27.4	14.8	18.2	4.7	2.3	2.4	2.9	1.4	1.4

Note: Long duration schemes are those schemes that are of duration greater than 30 months, and, similarly, short duration schemes are those which are below this benchmark.

Table 11. Default Rates by Monthly Contribution.

By Member (%)						By Total Transactions (%)					
High Value			Low Value			High Value			Low Value		
Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning	Overall	Before Winning	After Winning
31.3	18.5	20.0	34.3	20.0	20.0	3.1	1.5	1.6	4.3	2.1	2.2

Note: High value schemes in terms of monthly contribution are those schemes that have monthly contributions greater than Rs.3,000. Similarly, those schemes with lesser monthly contribution than this benchmark are low value schemes.

Table 12. Default Rates by Monthly Contribution-Chit Duration Combination.

By Member (%)			By Total Transactions (%)		
True High Value			True High Value		
Overall	Before Winning	After Winning	Overall	Before Winning	After Winning
28.6	16.2	18.3	2.8	1.4	1.4

Note: True high value in terms of monthly contribution and chit duration implies those schemes which are of high monthly contribution and short duration.

value schemes, as seen in Table 8. There is a significant drop in the percentage of members defaulting at least once and the percentage of total default transactions also exhibits a similar trend. This might imply that individuals seeking short duration high value loans are less risky than those seeking high value long duration loans.

One observation recorded while interviewing branch managers and company owners during the course of the field study was

that late payments and defaults increase in frequency around festivals. Table 13 shows no such seasonality in late payments, but defaults, especially defaults after taking prized amount, show a mild spike around the months of October and November, that is, around the time of the Indian festival of Diwali.

3.5. Collateral and surety information

The next table (Table 14) outlines the summary statistics of the collateral and surety information collected by the

Table 13. Seasonality in Late Payments and Defaults.

Month	Overall		Before Winning Prized Amount		After Winning Prized Amount	
	Total Late Payments	Total Defaults	Total Late Payments	Total Defaults	Total Late Payments	Total Defaults
January	4,570	1,073	1,898	505	2,672	568
February	4,407	1,082	1,818	519	2,589	563
March	4,268	1,002	1,867	488	2,401	514
April	4,470	1,019	1,881	527	2,589	492
May	4,392	1,065	1,861	541	2,531	524
June	4,334	1,060	1,838	530	2,496	530
July	4,228	1,132	1,801	551	2,427	581
August	4,177	1,139	1,819	549	2,358	590
September	4,381	1,065	1,863	518	2,518	547
October	4,492	1,118	1,862	540	2,630	578
November	4,482	1,191	1,804	572	2,678	619
December	4,465	1,098	1,871	517	2,594	581
<i>Average</i>	4,388.8	1,087.0	1848.6	529.8	2540.3	557.3

Table 14. Collateral and Surety Information.

Total Number of Chit Tickets	9,318
Total Number of Chit Tickets for which Collateral was provided	3,168
Types of Collaterals (%)	
Other Chits	59.5
Savings Certificates	4.9
LIC	17.6
Fixed Deposit Receipts	0.1
Land/Property Documents	17.9
Total Number of Chit Tickets for which Sureties were provided	4,481
Total Number of Sureties Provided	14,480
Average Number of Sureties	3.2
Minimum Number of Sureties	1
Maximum Number of Sureties Provided	6

companies. As stated in the section on the functioning and data collection process of the companies, collaterals and sureties are collected from individuals when they borrow from the companies (that is, when they win the prized amount and lay claim to the same). However, as stated earlier, a standard benchmark does not exist for the types of collateral and surety collected and all decisions on this matter are based on human judgement, thereby bringing in an inherent bias into the system.

In the dataset being analysed, we have information pertaining to a total of 9,318 chit tickets from 274 chit schemes. This figure excludes the tickets held by the companies themselves, and all these tickets are held by the 4,936 members being assessed. As is evident above, the companies have collected collateral or surety from approximately 82 per cent of the chit tickets. The remaining 18 per cent of the members are those who claim the prized amount at the end of the

scheme and, therefore, are not required to provide any collateral or surety. These are the individuals who are the pure savers in the schemes.

What is of interest, however, is the wide range of sureties that are collected—from one to six, with a mean of 3.2 sureties. This further indicates the need for a standardized procedure in assessing risk, because such a wide range of sureties implies the existence of biases towards certain types of members.

Another point of interest in this table is the type of collaterals provided by the members. A staggering 60 per cent of collaterals provided were other chits that the members were participating in. If we compare this figure, with the figures from [Table 5](#) and [Table 6](#), it implies that a large proportion of repeating members might be indulging in strategic financial planning using chit schemes, that is, participating in multiple schemes at the same time, using one scheme as a collateral for the other, and timing the borrowings and savings such that overall interest rates faced by the members would be minimized.

3.6. Auction information

The final tables attempt to describe the trends observed in the monthly auction process. As outlined in the earlier section, every month a competitive bidding process is carried out to determine which of the members of a given scheme collects the prized amount as a loan. The minutes of the said auctions are recorded by the companies, and this provides information on how often a member participates in auctions before winning ([Table 15](#)).

As can be seen in [Table 15](#), the average member participates in about four auctions before being able to claim the prized amount

as loan. The maximum number of attempts before a member can claim the prized amount appears to be 25. This, therefore, implies that there is a great unmet demand for liquidity. In this regard, [Table 16](#) outlines this fact in the example of the most frequently used chit scheme type, those of 30 months duration¹¹.

As can be observed, leaving aside the first auction, in the consequent first quarter of the scheme, the first 10 auctions, all face unmet demand for liquidity, or the percentage of unsuccessful bids, at rates greater than 50 per cent. A deeper concern is that in this period of the chit scheme, all participants bid at the maximum permissible cap, and winners of the prized amount are selected by a lottery mechanism, thereby increasing the randomness of meeting the demand for liquidity.

The rates then maintain an average of approximately 35 per cent till the last quarter of the scheme, and thereafter these rates drop. The drop in the last quarter corresponds to the fact that most of the members at that period are savers and therefore

¹¹For other chit scheme types, that is, for 20, 24, 25, 40, and 50 months, tables are available in tables A4, A5, A6, A7, and A8 in [Supplementary File 1](#).

Table 15. Participation in Auctions.

Average number of auctions participated in	4.2
Standard deviation	4.2
Minimum number of auctions participated in	1
Maximum number of auctions participated in	25

Table 16. Unmet demand for liquidity in 30 month schemes.

Auction Number	No. of Participants	No. of Winners	%age of Unsuccessful Bids	Auction Number	No. of Participants	No. of Winners	%age of Unsuccessful Bids
1	132	79	40.2	16	119	79	33.6
2	591	79	86.6	17	110	79	28.2
3	582	79	86.4	18	113	79	30.1
4	533	79	85.2	19	111	79	28.8
5	460	79	82.8	20	125	79	36.8
6	359	79	78.0	21	105	79	24.8
7	241	79	67.2	22	115	79	31.3
8	181	79	56.4	23	112	79	29.5
9	147	79	46.3	24	117	79	32.5
10	125	79	36.8	25	98	79	19.4
11	121	79	34.7	26	93	79	15.1
12	122	79	35.2	27	86	79	8.1
13	104	79	24.0	28	79	79	0.0
14	121	79	34.7	29	79	79	0.0
15	114	79	30.7	30	79	79	0.0

have no demand for liquidity. What is of interest to our study is the high demand in the first half of the scheme.

In summary, it appears that though default rates are overall low, there are some trends that need further analysis, for example, the effect of demographics on irregular payment behaviour, the effect of various trends in savings behaviour on the irregular payments, or the effect of high rates of unsuccessful bidding on the trends observed in irregular payments. However, it is obvious that a systemic overhaul is required to address the nature of risks faced by the chit fund industry. In the next sections we shall take a more in-depth look at these trends and formulate a methodology for analysing, assessing, and quantifying the risk faced by the companies.

4. Credit scoring model

The descriptive statistics of the dataset, as outlined in the previous section, gave us a birds-eye view of the underlying trends in the various types of information collected, and therefore acted as a guideline for the next round of in-depth analyses to identify the determinants of risks faced by the company and suggest an algorithm to mitigate the same.

In this section, we discuss the motivation for the in-depth analyses that were conducted and describe the methodologies followed.

4.1. Where does risk need to be assessed?

The role of an average chit fund company is centred around performing seven primary tasks: (1) locating potential members; (2) collecting and processing their information at the time of joining; (3) identifying a scheme where said member would fit in; (4) collecting monthly dues; (5) conducting auctions and identifying winner; (6) assessing level of surety and/or collateral applicable for winner; and (7) disbursing loan amount.

As we have already seen in Section II, the existing procedure of identifying risky clients followed by the chit fund industry

rallies around the collection of, what usually proves to be incomplete information at Point (2). Further on in the scheme, at Point (6), the assessment of risk and therefore the levels of surety or collateral required are again based on the information provided at Point (2), consequently adding to the risks already accrued to the companies at Point (2).

Therefore, standardized statistical methods need to enter the typical chit scheme at two points—at the point of collecting and processing demographic information, and at the point where assessment of the level of surety/collateral takes place.

At the first juncture, having a better knowledge about the new member is crucial. Demographic information, like age, income, gender, and type of occupation is collected now by the companies. This information, coupled with historical information about the member, like whether they were members prior to the current scheme that they are joining, whether they had defaulted in their earlier schemes, how many years have they been part of the company, etc., can improve the quality of assessing risk of the member.

Similarly, at the second juncture, when the member has been participating in a given chit scheme for a certain period of time, and takes part in an auction, wins the auction, and is about to claim the prized amount as a loan, another assessment needs to be carried out. At this juncture, not only will demographic information be available, but also the transaction history for the months that he/she has been participating for. Therefore, an assessment methodology that supplements decision-making processes by including the transaction history in the loop is needed.

4.2. Risk assessment at the point of entry

Due to the relationship-based screening of prospective clients, overall, the chit fund industry lacks a systematic collection of data at the point of entry. Clients are asked for basic demographic data only, and in many cases, based on the social network by which the client has been introduced to the company, even

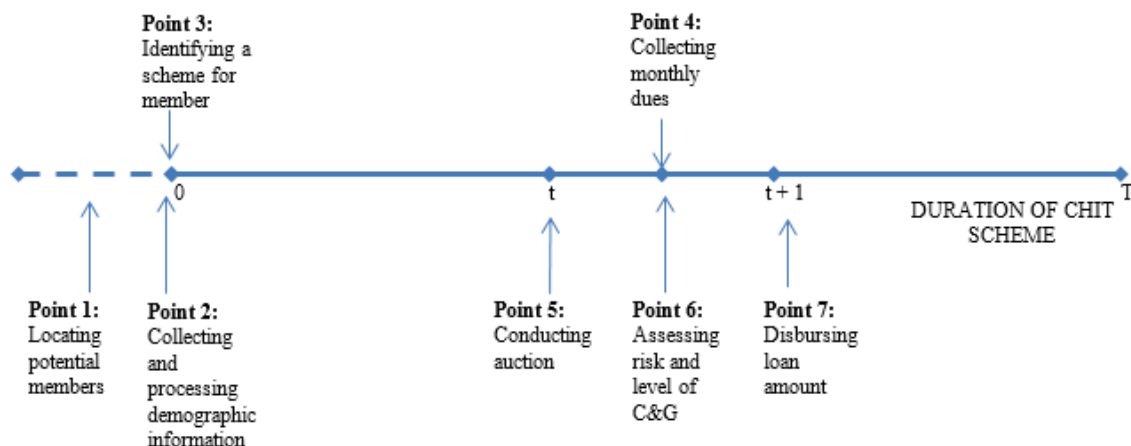


Figure 2. Tasks of a typical chit fund company.

these parameters are skipped. In general, for new members, only name, age, gender, income, and occupation information is collected. In the case of old members, the companies gloss over the updating of basic demographic parameters and only state the earlier scheme(s) the member had participated in. In both cases, characteristic biases are built into the data collection process, resulting in asymmetric information for the companies: (1) lack of prior financial history from prospective new clients does not allow the companies to judge the client's retrospective financial repayment behaviour; and (2) lack of updated demographic characteristics in the case of returning clients does not allow the companies to update existing risk profiles of the clients. Furthermore, in the case of Point (1) above, there is no information sharing mechanism that exists across chit fund companies, resulting in clients being able to hide irregularities within the industry.

Therefore, all risk assessment at this juncture needs to be based on the following parameters: (1) stated demographic indicators, that is, age, gender, occupation, and income; (2) whether the client participated in schemes earlier with the same company; and (3) whether the client displayed any irregular payment behaviour in the schemes he/she participated in earlier.

Keeping the data availability in mind, we propose a model which allows us to analyse the above stated parameters in order to identify specific determinants of increasing total defaults. Thereafter, using the identified determinants, develop a rule by which prospective clients can be rated on a scale of Low Risk to High Risk.

The dataset used for this analysis is cross-sectional—each member's demographic characteristics and prior repayment records are aggregated member-wise.

4.3. Risk assessment at the point of claiming prized amount as loan

At the point where a client wins the auction process and lays claim to the prized amount as a loan, he/she officially accrues credit risk to the company by formally becoming a borrower. This point can also be called an inflection point in the characteristics of the client—he/she changes from being an apparent saver to being a borrower. Therefore, an assessment of risk accrued to the company by the client at this juncture is logical.

Having said that, an analysis of available data at this juncture is crucial; it can be safely assumed that the data available to the company has been greatly supplemented since the time the client joined the scheme. At the time of joining, the company has basic demographic information about the client, that is, age, gender, occupation, and income level. In addition to this, the company also maintains record of whether the client has been an earlier member of the company or not. Adding to the available data is: (1) transaction records of the individuals over the months that they have been participating in

the scheme; and (2) retrospective non-prized bidding data, that is, information pertaining to the individual's demand for loans.

The conglomerated data available for each member at the point of their taking a loan allowed us to consult existing models of credit scoring available in the formal banking industry. For example, the Fair Isaac Corporation (FICO) methodology of consumer credit scoring that is accepted as the industry standard across the world.

FICO scores are widely-used by banks around the world to assess the credit worthiness of clients. In many countries, banks are not allowed to deny loans to applicants, but are allowed to raise interest rates for clients who are assessed to expose the bank to higher risk. In the Indian context, banks and formal non-banking financial organizations use these scores to determine credit eligibility by assessing the risk level of a prospective borrower.

Since the exact scoring methodology is the intellectual property of FICO, and the exact formula is unavailable in the public domain, a break-up of the weights given to different categories of a client's financial history is available and is listed in [Table 17](#).

In comparing the above category heads with the type of information available in the dataset at the point of an individual laying claim to the prized amount as loan, we were able to identify variables that could be used to build a credit scoring model loosely based on the FICO methodology. This choice of said variables that are applicable to the chit fund industry, are listed in the final column.

The above variables, in addition to the demographic and prior records of participation in chit schemes of the members, were arranged in an unbalanced panel format. That is, for each member, we have data for each month he/she participated in the scheme up to the point of him/her taking a loan.

For example, in the case where risk assessment was done at the time of a member joining a scheme, the identification and analysis of determinants of risky behaviour takes place again, followed by the development of a relative score that rates members from being Low Risk to High Risk.

5. Regression results

As discussed in the previous section, there are two points in the duration of a member's participation in a given chit scheme where risk assessment should be carried out by the company: (1) at the time of a member joining a scheme; and (2) at the time the member withdraws the prized amount as a loan. In this section, we shall discuss the regression results corresponding to the above two points in a member's participation time line, and outline how the results were used to build a relative scorecard for tagging risky members.

Table 17. FICO Model and Application to Chit Fund Industry.

FICO Category Head	Various Aspects under given Category	Methodology for use in our Analysis
Payment History	Types of accounts held	Not applicable, since we do not have data for out of CF company financial history
	Account payment information for each type of account held	We are tracking types of payments made for CF accounts
	Presence of adverse public records, delinquent payments	We do not have records of financial history out of CF company, but delinquent payments within company is tracked
	Severity of delinquency	We are considering the maximum duration of delinquency
	Amounts due on delinquent accounts	We are considering the amount due as a fraction of respective chit value for each transaction made
	Time since last occurrence of adverse public record or delinquent payment	We are considering duration since last delinquent payment
	Number of current delinquent items	We are considering the number of times a client makes delinquent payments
	Number of accounts with no delinquency records	We are using number of simultaneous schemes with no defaults
Amounts Owed	Total amount owed	We have variables that account for future liability of a client
	Total amount owed by different accounts	Not used
	Lack of specific type of balance in various type of accounts	We are considering variables like total amount due, total amount paid, comparison between future liability and loan amount, etc.
	Number of accounts with no dues	Not used
	Proportion of balances to total credit limits (credit flow information)	We are considering future liability to total instalments paid
	Proportion of outstanding dues to current balances	
Length of Credit History	Time since accounts opened for various types of accounts	We have variables that account for the number of years a client has been a member of a company
	Activity in accounts	We have variables that account for total active period in company for a given client
New Credit	Number of recently opened accounts	Not used
	Proportions of balances to dues in each new type of recently opened accounts	
	Number of recent credit inquiries	We have considered the duration since last participation in auction and number of lost bids
	Time since last credit inquiry	
Type of Credit Used	Number of types of credit accounts held	Not applicable

5.1. Regression analyses at the time of joining

The dataset used for this set of regression analyses was cross-sectional in nature, where each member's details were aggregated to form one line of information. The analysis conducted was retrospective in nature. For each member, his/her age, gender, occupation, income level, and prior participation in terms of number of schemes he/she has completed in the given company were considered.

In this analysis, we wanted to study the determinants of defaults at the member level. Therefore, we intended to use the total number of defaults as the dependent variable, and explain what drives defaults in terms of demographic variables. That is,

$$\text{Total Defaults} = F(\text{Demographic variables})$$

Thus, we aggregated the total number of defaults each member had committed in all the schemes they participated in. As we can see from the descriptive statistics of irregular payments, approximately 35 per cent of members have defaulted at least once in their participatory history¹². Therefore, in our sample space of 4,936 members, a large number of members had zero values for total number of defaults, and among the 35 per cent who had non-zero defaults there appeared to be a large

¹²Refer Table 7.

skewness. To counter this, we decided to use a logarithmic transform of the total defaults variable, and consequently the equation for estimation became as follows:

$$\text{Log}(1 + \text{total defaults}) = F(\text{Demographic variables})$$

Similarly, logarithmic transforms of total defaults before winning prized amount and total defaults after winning prized amount were also created.

To counter the effects of skewness in the demographic variables, we created a logarithmic transform of age, and categorized income into high, medium, and low levels, with the reference group being all members with missing income levels. The quadratic form of age was also considered. In a similar manner, dummy variables were created to represent male members, whereby all male members took the value 1, and female members were the reference group; and, all company tickets took the value 1, and female members were again the reference group. In case of occupation categories, each occupation group was represented by a dummy variable, with 'other occupations' being the reference group in each case¹³.

Following this process, we created categories based on occupation and income levels, as outlined in [Table 18](#).

The final regression models that were estimated were as follows:

$$\text{Log}(1 + \text{Total Defaults}) = F[\text{Log}(\text{age}), (\text{Log}(\text{age}))^2, \text{Number of schemes}, \text{Dummy}(\text{male}), \text{Dummy}(\text{company}), \text{var } 1 \dots, \text{var } 21]$$

$$\text{Log}(1 + \text{Total Defaults, before winning}) = F[\text{Log}(\text{age}), (\text{Log}(\text{age}))^2, \text{Number of schemes}, \text{Dummy}(\text{male}), \text{Dummy}(\text{company}), \text{var } 1 \dots, \text{var } 21]$$

$$\text{Log}(1 + \text{Total Defaults, after winning}) = F[\text{Log}(\text{age}), (\text{Log}(\text{age}))^2, \text{Number of schemes}, \text{Dummy}(\text{male}), \text{Dummy}(\text{company}), \text{var } 1 \dots, \text{var } 21]$$

The results estimated from the above models are shown in [Table 19](#).

¹³Refer [Table 4](#).

Table 18. Occupation and Income Level Categories and Variables.

Occupation Income Level	Low Income	Medium Income	High Income
Business Owner	Var 1	Var 2	Var 3
Government Service	Var 4	Var 5	Var 6
Housewife	Var 7	Var 8	Var 9
Self-employed Professional	Var 10	Var 11	Var 12
Private Sector Service	Var 13	Var 14	Var 15
Retired	Var 16	Var 17	Var 18
Small Business Owner	Var 19	Var 20	Var 21

As seen in the above table, when attempting to identify the determinants of increasing total defaults at the time of joining, the following trends were noticed.

If a member participates in an additional scheme, his/her tendency to default at least once during the duration of a scheme increases by 12 per cent. This implies that when a person has been a member of a company for a longer duration of time, the greater the chance of him/her defaulting, indicating that there exists a significant relaxation of stringency of rules being applied to the member by the companies—a direct off-shoot of the relationship maintained by companies and their clients.

Demographic parameters—age and gender appear to have no significant effect on total defaults. However, it appears that those individuals who mention their income and occupation tend to default lesser than those who do not, as evident from the negative deviations from mean.

When looking at categories of income and occupation, the following groups appear to deviate, in a negative manner, significantly from the total defaults:

1. Self-employed professionals with high income
2. Government sector workers with high income
3. Government sector workers with medium income
4. Government sector workers with low income
5. Housewives with high income
6. Private sector employees with high income
7. Private sector employees with medium income
8. Private sector employees with low income
9. Housewives with medium income
10. Self-employed professionals with low income
11. Self-employed professionals with medium income
12. Housewives with low income
13. Small business owners with high income
14. Small business owners with low income
15. Small business owners with medium income
16. Retired individuals with low income

The only parameter that affects a member's defaults before winning a prized amount is greater participation in schemes; each additional scheme tends to increase total defaults before winning by 10 per cent.

When looking at the parameters affecting total defaults after winning a prized amount, the following trends were observed.

If a member participates in an additional scheme, there is an increase in total defaults after winning of 5 per cent. As

Table 19. Regression part I: OLS (Ordinary Least Squares) regression at the time a client joins the company.

Independent Variables	Log(1+ total defaults)	Log(1+ total defaults, before winning)	Log(1+ total defaults, after winning)
Number of chit schemes participated in	0.120*** (11.03)	0.099*** (10.32)	0.0501*** (5.29)
Log of age	0.774 (0.78)	-0.968 (-1.16)	2.301*** (3.35)
Log of age, squared	-0.141 (-1.03)	0.141 (1.23)	-0.363*** (-3.90)
Dummy (male = 1, female = 0)	0.045 (1.62)	-0.026 (-1.23)	0.0697** (3.29)
Dummy (company ticket = 1, female = 0)	-0.139 (-0.39)	-0.209 (-0.66)	-0.0356 (-0.14)
Business Owner, Low Income	-0.348 (-1.70)	-0.177 (-1.13)	-0.173 (-1.09)
Business Owner, Medium Income	-0.211 (-1.09)	-0.060 (-0.40)	-0.130 (-0.86)
Business Owner, High Income	-0.224 (-1.22)	0.018 (0.12)	-0.223 (-1.51)
Government Service, Low Income	-1.304* (-2.54)	-0.256 (-0.69)	-1.149** (-2.63)
Government Service, Medium Income	-1.321* (-2.52)	-0.422 (-1.13)	-0.976* (-2.17)
Government Service, High Income	-1.344* (-2.57)	-0.548 (-1.49)	-0.888* (-1.96)
Housewife, Low Income	-1.015* (-2.47)	-0.300 (-0.99)	-0.743* (-2.18)
Housewife, Medium Income	-1.045* (-2.44)	-0.319 (-1.00)	-0.650 (-1.86)
Housewife, High Income	-1.159** (-2.72)	-0.369 (-1.14)	-0.795* (-2.38)
Self Employed Professional, Low Income	-1.035* (-2.37)	-0.202 (-0.60)	-0.932** (-2.64)
Self Employed Professional, Medium Income	-1.032* (-2.50)	-0.193 (-0.64)	-0.910** (-2.59)
Self Employed Professional, High Income	-1.381*** (-3.54)	-0.467 (-1.71)	-1.049** (-3.05)
Private Sector, Low Income	-1.049** (-2.64)	-0.323 (-1.11)	-0.792* (-2.37)
Private Sector, Medium Income	-1.094** (-2.64)	-0.386 (-1.28)	-0.757* (-2.17)
Private Sector, High Income	-1.159** (-2.84)	-0.415 (-1.40)	-0.800* (-2.34)
Retired, Low Income	-0.560* (-2.14)	-0.269 (-1.29)	-0.287 (-1.48)
Retired, Medium Income	-0.530 (-1.82)	-0.193 (-0.81)	-0.322 (-1.56)

Independent Variables	Log(1+ total defaults)	Log(1+ total defaults, before winning)	Log(1+ total defaults, after winning)
Retired, High Income	-0.476 (-1.49)	-0.130 (-0.51)	-0.262 (-1.16)
Small Business Owner, Low Income	-0.890* (-2.57)	-0.278 (-1.10)	-0.652* (-2.26)
Small Business Owner, Medium Income	-0.849* (-2.40)	-0.221 (-0.84)	-0.649* (-2.24)
Small Business Owner, High Income	-0.998** (-2.92)	-0.443 (-1.86)	-0.613* (-2.11)
Probability of choosing high value scheme	0.353 (0.24)	-0.531 (-0.47)	1.698 (1.48)
Probability of choosing high monthly contribution schemes	0.0192 (0.02)	0.428 (0.52)	-0.418 (-0.48)
Probability of choosing long duration schemes	8.153* (2.29)	1.538 (0.62)	7.228* (2.27)
Probability of choosing true high value schemes	-1.054 (-0.63)	-0.0198 (-0.01)	-2.218 (-1.72)
Intercept	-5.221* (-2.30)	0.955 (0.55)	-7.471*** (-4.16)
Total Number of Observations	3698	3698	3698
R ²	0.124	0.142	0.063

Notes: (1) Definition of true high value schemes remains the same as outlined in the Summary Statistics tables, that is, those schemes that are of high monthly contribution and short duration; (2) t-values are stated in parentheses below the parameter estimates; and (3) *** indicates confidence of 99 per cent, ** indicates confidence of 95 per cent, and * indicates confidence of 90 per cent.

mentioned earlier, this correlates with the possible relaxation of rules with older members.

Interestingly, demographic effects have a greater significance here: an increase in the age of the member, results in an increase in the total defaults after winning at a decreasing rate; male members tend to have a 6% higher rate of defaults after winning than women; and again, it appears that those individuals who mention their income and occupation tend to have a lower tendency to default, as seen in the negative deviations from the reference group (those individuals who did not mention their income and occupation).

When looking at categories of income and occupation, the following groups appear to deviate, in a negative manner, significantly in the total defaults:

1. Government sector worker with low income
2. Self-employed professional with high income
3. Government sector worker with medium income
4. Self-employed professional with low income
5. Self-employed professional with medium income
6. Government sector worker with high income
7. Private sector employee with high income

8. Housewives with high income
9. Private sector employee with low income
10. Private sector employee with medium income
11. Housewives with low income
12. Small business owners with low income
13. Small business owners with medium income
14. Small business owners with high income

5.2. Derivation of credit score at the time of joining

Using the results from above, a few deductions about riskiness of clients can be made directly: (1) the length of time spent with a company should not be used as a determinant of stringency of rules applicable to the member, since it is apparent that these type of members display a statistically significant increase in total number of defaults; (2) youngest and oldest members are the least risky, and middle-aged members show a large tendency to default; and (3) male members are riskier than women members.

In order to quantify the riskiness observed in the above regressions, we decided to categorize individuals into various risk types based on their income and occupation categories. For achieving this, we considered the statistically significant deviation

of the corresponding category dummy from the mean, that is, assuming the reference group to be all individuals who had not stated their income and occupation at the time of joining; we consider the deviation of those who did mention income and occupation. In doing so, we observe that all groups have a negative deviation from the mean, and we proceed to rank all income and occupation categories by these deviations, as seen in [Table 20](#). Maximum deviation implies least risk, and no deviation implies maximum risk. The switchover between each risk type and score is based on whether the difference between each consecutive estimate is significantly greater than zero or not.

From the above ranking, we see that Self-employed Professionals with high income are the least risky type of members, therefore given a risk type of 'A' and a score of 1. Similarly, moving down the table, the riskiness increases till Rank 17, where a cluster of income and occupation categories display zero deviation from mean, and are given the risk type 'E' and a score of 5.

The same methodology is applied to the regression results obtained from the third model—to identify the determinants of total regressions after winning the prized amount. The risk types and scores thus obtained are outlined in [Table 21](#).

From the above ranking it appears that those members with government sector jobs and those in the lowest income category are the least risky, with a risk type 'A' and a score of 1. Business owners, retired individuals, and housewives with medium income level are the most risky, with a risk type of 'E' and score of 5.

On combining the above two tables, we are able to derive a score for total riskiness of clients at the time of joining and this is shown in [Table 22](#).

As per the above table, the final scores derived at the time of joining indicate that the least risky groups are government servants with low income and self-employed professionals with high income. The groups with maximum risk are business owners and retired individuals with medium and high income.

Table 20. Scores derived for Overall Riskiness.

Rank	Occupation/Income Categories	Relative Deviation from Mean	Risk Type	Score
1	Self Employed Professional, High Income	-1.381	A	1
2	Government Service, High Income	-1.344	A	1
3	Government Service, Medium Income	-1.321	A	1
4	Government Service, Low Income	-1.304	A	1
5	Housewife, High Income	-1.159	B	2
6	Private Sector, High Income	-1.159	B	2
7	Private Sector, Medium Income	-1.094	B	2
8	Private Sector, Low Income	-1.049	B	2
9	Housewife, Medium Income	-1.045	B	2
10	Self Employed Professional, Low Income	-1.035	C	3
11	Self Employed Professional, Medium Income	-1.032	C	3
12	Housewife, Low Income	-1.015	C	3
13	Small Business Owner, High Income	-0.998	C	3
14	Small Business Owner, Low Income	-0.89	D	4
15	Small Business Owner, Medium Income	-0.849	D	4
16	Retired, Low Income	-0.56	D	4
17	Business Owner, Low Income	0*	E	5
	Business Owner, Medium Income	0*	E	5
	Business Owner, High Income	0*	E	5
	Retired, Medium Income	0*	E	5
	Retired, High Income	0*	E	5

Note: (*) Relative deviation of 0 implies that for the corresponding category, the deviation from mean was statistically insignificant.

Table 21. Scores Derived for Riskiness after Winning Prized Amount.

Rank	Occupation/Income Categories	Relative Deviation from Mean	Risk Type	Score
1	Government Service, Low Income	-1.149	A	1
2	Self Employed Professional, High Income	-1.049	A	1
3	Government Service, Medium Income	-0.976	B	2
4	Self Employed Professional, Low Income	-0.932	B	2
5	Self Employed Professional, Medium Income	-0.910	B	2
6	Government Service, High Income	-0.888	B	2
7	Private Sector, High Income	-0.800	C	3
8	Housewife, High Income	-0.795	C	3
9	Private Sector, Low Income	-0.792	C	3
10	Private Sector, Medium Income	-0.757	C	3
11	Housewife, Low Income	-0.743	D	4
12	Small Business Owner, Low Income	-0.652	D	4
13	Small Business Owner, Medium Income	-0.649	D	4
14	Small Business Owner, High Income	-0.613	D	4
15	Business Owner, Low Income	0*	E	5
	Business Owner, Medium Income	0*	E	5
	Business Owner, High Income	0*	E	5
	Housewife, Medium Income	0*	E	5
	Retired, Low Income	0*	E	5
	Retired, Medium Income	0*	E	5
	Retired, High Income	0*	E	5

Note: (*) Relative deviation of 0 implies that for the corresponding category, the deviation from mean was statistically insignificant.

Table 22. Final Scores at the Time of Joining.

S. No.	Occupation/Income Category	Score 1 (5)	Score 2 (5)	TOTAL SCORE AT TIME OF JOINING (10)
1	Government Service, Low Income	1	1	2
2	Self Employed Professional, High Income	1	1	2
3	Government Service, Medium Income	1	2	3
4	Government Service, High Income	1	2	3
5	Housewife, High Income	2	3	5
6	Private Sector, Low Income	2	3	5
7	Private Sector, Medium Income	2	3	5
8	Private Sector, High Income	2	3	5
9	Self Employed Professional, Low Income	3	2	5
10	Self Employed Professional, Medium Income	3	2	5
11	Housewife, Low Income	3	3	6
12	Housewife, Medium Income	2	5	7
13	Small Business Owner, High Income	3	4	7
14	Small Business Owner, Low Income	4	4	8
15	Small Business Owner, Medium Income	4	4	8
16	Retired, Low Income	4	5	9
17	Business Owner, Low Income	5	5	10
18	Business Owner, Medium Income	5	5	10
19	Business Owner, High Income	5	5	10
20	Retired, Medium Income	5	5	10
21	Retired, High Income	5	5	10

5.3. Regression analysis at the time of taking loan

The dataset used for this set of regression analyses was an unbalanced panel, where each member's monthly transaction details were tracked till the month the member claimed the prized amount as loan. This allowed us to assess the savings behaviour of the individual until he/she becomes a borrower, and use observable transactional behaviour characteristics in identifying the determinants of defaulting as a borrower. As outlined in the previous section, apart from the demographic characteristics, each member's transaction history will be used to assess the member's records of delinquency, severity of delinquency, demand for loans, outstanding dues, etc., and all these, in turn, will be used to determine the probability of a member defaulting as a borrower.

For identifying the determinants of defaults in this analysis, probit models of the following form were defined and estimated:

$$\text{Pr}(\text{Default after winning} = 1) = F[\text{transaction history as a saver, demographic characteristics}]$$

As in the earlier regression models, to counter the skewness in demographic variables, a logarithmic transformation of age was considered, dummy variables for male members and company members were included, and, finally, categories for income and occupation were created¹⁴.

The results estimated from the above models are listed in Table 23.

From the above table, some clear trends related to transaction history were observed. First, with respect to the choice of schemes, participation in longer duration and high monthly

¹⁴Refer Table 18.

Table 23. Regression Part II: Probit Regression at the Time of Borrowing.

Independent Variables	Pr(Default after winning = 1)	Pr(Default after winning = 1)
Log of Duration of Particular Scheme	2.646*** (47.33)	2.399*** (39.97)
Log of Monthly Contribution of Particular Scheme	2.399*** (46.21)	2.183*** (39.12)
Log of Prized Amount (Loan Amount)	-2.445*** (-47.78)	-2.246*** (-40.90)
Log of Years of Activity in Company	-0.332*** (-22.26)	-0.263*** (-15.38)
Percentage of Scheme Completed	-0.306*** (-8.63)	-0.322*** (-8.37)
Proportion of outstanding dues to loan amount	-0.00742 (-1.55)	-0.00670 (-1.28)
Proportion of early transactions to total transactions made	-0.114 (-1.69)	0.00134 (0.02)
Proportion of part transactions to total transactions made	-0.196*** (-3.65)	-0.265*** (-4.49)
Proportion of late transactions to total transactions made	0.240** (2.91)	0.359*** (4.01)
Proportion of default transactions to total transactions made	-1.129*** (-11.61)	-1.267*** (-11.44)
Proportion of lump-sum transactions to total transactions made	0.659*** (12.70)	0.627*** (11.07)
Maximum duration of default	0.0837*** (37.54)	0.0854*** (30.54)
Proportion of lost bids to total auctions held	0.443*** (20.19)	0.497*** (20.99)
Duration from last missed instalment	-0.0564*** (-26.03)	-0.0553*** (-23.63)
Proportion of difference of amount due and amount paid by member and chit value	3.820*** (22.32)	4.414*** (22.00)
Total number of other payments (for other schemes) being made at that juncture	0.0279*** (8.98)	0.00966** (2.68)
Log of Age		6.996*** (14.11)

Independent Variables	Pr(Default after winning = 1)	Pr(Default after winning = 1)
Log of Age, squared		-0.998*** (-14.69)
Dummy(=1 if male, =0 otherwise)		0.120*** (8.05)
Business Owner, Low Income		-0.131*** (-3.52)
Business Owner, Medium Income		0.0287 (0.96)
Business Owner, High Income		0.198*** (4.84)
Government Service, Low Income		-1.555*** (-15.54)
Government Service, Medium Income		-0.365*** (-8.67)
Government Service, High Income		-0.209*** (-3.30)
Housewife, Low Income		-0.397*** (-6.53)
Housewife, Medium Income		0.0108 (0.16)
Housewife, High Income		0.0338 (0.36)
Self Employed Professional, Low Income		0.474*** (4.46)
Self Employed Professional, Medium Income		0.0705 (1.00)
Self Employed Professional, High Income		(dropped)*
Private Sector, Low Income		-0.371*** (-12.50)
Private Sector, Medium Income		-0.256*** (-12.24)
Private Sector, High Income		-0.298*** (-12.38)
Retired, Low Income		-0.0618 (-1.18)
Retired, Medium Income		-0.284*** (-4.78)
Retired, High Income		-0.970*** (-4.52)
Small Business Owner, Low Income		-0.539*** (-9.41)
Small Business Owner, Medium Income		-0.432*** (-7.55)
Small Business Owner, High Income		-0.165 (-1.94)
Intercept		-13.34*** (-14.50)
Total Number of Observations	90,152	79,972
R ²	0.222	0.221

Note: (1) (+) The estimate for self employed professional with high income was dropped since it appeared to perfectly predicts no defaults; (2) t-values are stated in parentheses below the parameter estimates; (3) *** indicates confidence of 99 per cent, ** indicates confidence of 95 per cent and * indicates confidence of 90 per cent.

contribution schemes increases the probability of defaulting as a borrower. However, higher loan amounts (prized amounts) indicate a lower probability of defaulting.

As a member remains active in a company for a longer duration, he/she exhibits a lower probability of defaulting as a borrower. Comparing this result with what is obtained in the first set of regressions, we can infer that though an older member faces laxer rules, he/she exhibits risky behaviour only as a saver (time period prior to claiming prized amount)¹⁵.

If a member participates in a scheme for a longer duration before claiming a prized amount, he/she has a lower probability to default. This implies that if a member exhibits saving behaviour for a longer duration before taking a loan, he/she will have a lower probability of default.

In terms of payment behaviour, if a member increases the number of part (incomplete) payments, then he/she has a lower probability of default. Since part payments imply a member's willingness to pay his/her dues, this result implies that though a person might not be able to make complete payment of dues, his/her willingness to do the same shows a reduction in the probability of default.

If a person makes late payments during his/her period as a saver, he/she has a greater probability of defaulting as a borrower. This also gives credence to the fact that bad savings behaviour gives rise to bad borrowing behaviour.

If a member defaults as a saver, he/she tends to have a lower probability of defaulting as a borrower. Conversely, if a member's default period increases, his/her probability of defaulting as borrower increases.

Members making more lump-sum payments tend to have a greater probability of defaulting as a borrower.

Combining the trends observed in payment behaviour with the fact that being a saver for a longer duration implies lower probability of default, it can be deduced that good payment behaviour during the period a member is a saver, in other words good savings behaviour results in good payment behaviour during the period as a borrower.

Two other interesting observations from the transaction history related results are: (1) a member showing a greater desperation for a loan, that is, if a member participates unsuccessfully in a greater number of auctions, he/she will have a greater probability of defaulting; and (2) at any given point of time, if a member makes payments for additional schemes, then he/she has a greater probability of defaulting as a borrower.

With respect to demographics, trends similar to the regression results from the time of joining were observed. That is, as age

increases, probability of defaulting as a borrower increases at a decreasing rate; male members display a greater probability of defaulting as a borrower; and, those members reporting income and occupation tend to default lesser than those who do not.

When considering income and occupation categories, the following groups appear to have significantly higher probability of defaults as a borrower:

1. Government Service, Low Income
2. Retired, High Income
3. Small Business Owner, Low Income
4. Small Business Owner, Medium Income
5. Housewife, Low Income
6. Private Sector, Low Income
7. Government Service, Medium Income
8. Private Sector, High Income
9. Retired, Medium Income
10. Private Sector, Medium Income
11. Government Service, High Income
12. Business Owner, Low Income
13. Business Owner, High Income
14. Self Employed Professional, Low Income

5.4. Derivation of credit score at the time of borrowing

Using the results from above, a few immediate deductions can be made with regards to the riskiness of members at the time of borrowing. In other words, if a member displays, or has displayed these trends during the period of him/her being a saver, then he/she should be labelled as a risky client: (1) borrowing at the early stages of a scheme; (2) making a large number of late payments or lump-sum payments; (3) participating in auctions frequently; (4) participating in more than one scheme at a time; (5) being middle-aged; and (6) male.

Following this, we used the same methodology as that used at the time of joining, to categorize members into various risk types. Consequently, we were able to generate scores for the different categories of income and occupation as listed in [Table 24](#).

From the above ranking, we see that Self-employed professionals with high income continue to remain the least risky category in terms of income and occupation. Small business owners with high income, self-employed professionals with medium income, business owners with medium income, housewives with high and medium income, and retired individuals with low income are the riskiest type of members.

In combining the rankings obtained above, with the rankings obtained in [Table 22](#), we created a comprehensive score with

¹⁵Refer [Table 19](#).

Table 24. Scores Derived for Riskiness at Time of Borrowing.

Rank	Occupation/Income Categories	Relative Deviation from Mean	Risk Type	Score
0 ⁺	Self Employed Professional, High Income	–	–	0
1	Government Service, Low Income	–1.555	A	1
2	Retired, High Income	–0.970	A	1
3	Small Business Owner, Low Income	–0.539	B	2
4	Small Business Owner, Medium Income	–0.432	B	2
5	Housewife, Low Income	–0.397	B	2
6	Private Sector, Low Income	–0.371	B	2
7	Government Service, Medium Income	–0.365	B	2
8	Private Sector, High Income	–0.298	C	3
9	Retired, Medium Income	–0.284	C	3
10	Private Sector, Medium Income	–0.256	D	4
11	Government Service, High Income	–0.209	D	4
12	Business Owner, Low Income	–0.131	D	4
13	Business Owner, High Income	0.198	E	5
14	Self Employed Professional, Low Income	0.474	E	5
15	Small Business Owner, High Income	0*	E	5
	Self Employed Professional, Medium Income	0*	E	5
	Retired, Low Income	0*	E	5
	Housewife, High Income	0*	E	5
	Business Owner, Medium Income	0*	E	5
	Housewife, Medium Income	0*	E	5

Note: (+) the regression results in Table 23 suggest that no self-employed professional with high income default after borrowing, and, therefore, they are the least risky category; (*) Relative deviation of 0 implies that for the corresponding category, the deviation from mean was statistically insignificant.

which to rank clients. These scores, and the rankings thereof, are shown in Table 25.

As reflected in the Table 25, the least risky group appears to be Self-employed professionals with high income, and the riskiest group appears to be Business owners with medium and high income.

6. Application of findings to the industry

As discussed in the earlier sections, the credit scoring model developed here will be playing a complementary role to human judgement. In order to visualize the role of the credit scoring model, let us consider the life-cycle of a member from the time he/she is a prospective client till the time he/she completes their participation in a given scheme.

The first stage in the life-cycle of a member is when he/she is a prospective client and the company is persuading them to

join a chit fund scheme. At this point, the client would need to fill up an application form. As per our field studies, we have noticed that the basic information requirement for an application form are statements of income, occupation, age, gender, and records of prior participation in the same company. The first part of the credit scoring model would be applied at this point to rate the client, ex-ante, on the following: (1) whether he/she falls into an overall risky group by being an old member or participating in simultaneous schemes or being male; and (2) what his/her credit score is, based on their income and occupation categories.

This ex-ante rating would not bar the member from participating in a scheme, but increase the monitoring from the company-side.

Once the member has started participating in a scheme, every month he/she will be generating transactional data based on his/her interactions with the company. Based on the incoming

Table 25. Final Score at the Time of Borrowing.

S. No.	Occupation/Income Category	Score at Time of Joining (10) ¹⁶	Score at time of taking loan ¹⁷ (5)	Final Score (15)	Rank
1	Self Employed Professional, High Income	2	0	2	1
2	Government Service, Low Income	2	1	3	2
3	Government Service, Medium Income	3	2	5	3
4	Government Service, High Income	3	4	7	4
5	Private Sector, Low Income	5	2	7	4
6	Housewife, Low Income	6	2	8	6
7	Private Sector, High Income	5	3	8	6
8	Private Sector, Medium Income	5	4	9	8
9	Housewife, High Income	5	5	10	9
10	Self Employed Professional, Low Income	5	5	10	9
11	Self Employed Professional, Medium Income	5	5	10	9
12	Small Business Owner, Low Income	8	2	10	9
13	Small Business Owner, Medium Income	8	2	10	9
14	Retired, High Income	10	1	11	14
15	Housewife, Medium Income	7	5	12	15
16	Small Business Owner, High Income	7	5	12	15
17	Retired, Medium Income	10	3	13	17
18	Business Owner, Low Income	10	4	14	18
19	Retired, Low Income	9	5	14	18
20	Business Owner, Medium Income	10	5	15	20
21	Business Owner, High Income	10	5	15	20

data every month, the company can decrease their monitoring of the member if the member displays the following attributes: (1) delaying borrowing, or acting as a saver for a longer duration; and (2) maintaining an image of strong willingness to pay—lack of late payments, making part payments, or early payments. Similarly, the company can increase their monitoring of the member if the member does the following: (1) displays increased irregular payments—late payments, lump-sum payments or defaults as a saver; (2) increased participation in auctions; and (3) participates in multiple simultaneous schemes.

At the time the member borrows the prized amount, a second credit score is generated on the basis of income and occupation categories. This score, in conjunction with the ex-ante score determines how risky the member will be as a borrower. If the score is high, then the company can take the following steps to mitigate risk: (1) increase the collateral and surety requirements for the member; and (2) increase monitoring of the member and perhaps design a member-specific repayment mechanism.

Therefore, the credit scoring model developed here will not only act as an early warning system to tag risky clients, but also

act as a constant warning mechanism for clients as they make payments.

7. Scalability

When speaking of scaling up the credit scoring model developed in this project, we need to first consider that immediate implementation might be impossible given the vast disparity in the quality of data maintained by the companies. This is evident from the final dataset used by us: we collected data from five companies, but were finally able to use only the best quality data which amounted to two companies.

Therefore, the first step in scalability should be a fine-tuning of processes in the companies, starting with the data collection methodology—collecting as much data as mandated by the application form design of the companies. Currently we see that even though the companies ask for only age, gender, income, and occupation, many applications do not have even these fields completed. In our data collection phase, we had to access

¹⁶Refer Table 22.

¹⁷Refer Table 24.

this information from various sources—application forms, surety documents, photocopies of proofs of occupation and residence that were submitted, etc. As a first step, all this information should be maintained in one location, and preferably digitized for easy access.

Following this, the next step would entail application forms and surety forms being redesigned to collect more information about clients; for example, education details, durable assets, immovable assets, outstanding loans from other sources, details of workplace, and details of occupation. Furthermore, great care has to be taken to increase the frequency at which the data is updated. This will improve the companies' information database and, therefore, give the company a better understanding of their clients—an improvement in knowledge of clients will act as a first level of risk assessment since information asymmetry will be reduced.

The steps mentioned above need to be complemented with systematic investment in technology. All client records should be digitized to begin with. Following this, further investments in technology, corresponding to transactional functioning of the companies should be looked into; for example, in the fields of collection and disbursing of dues, monitoring of auctions, filing of sureties, etc. Short-term investment in technology will result in an overall reduction in operational costs for the company and, therefore, allow them to expand their target population coverage.

Only after a standardization in the processes is achieved can a systematic credit scoring platform be implemented in a phased manner across the industry. The first target should be mid-sized firms who appear to be on the threshold of expansion, but are held back due to high operational ratios and high perceived risks associated with new clients; standardizing their risk assessment and mitigation methodologies, and fine-tuning their technological investments should help address these issues.

Following mid-sized firms, small-sized firms can be targeted. These firms are currently reliant heavily on their relationship lending model and are unlikely to take any change favourably. The ability to implement technological investments and credit scoring platforms would be based on their future growth path as a company. And, finally, large-sized firms could be targeted. However, these firms already have a set system in place and changing their system might not be optimal, and only a fine-tuning of their methodologies with our expertise might be possible.

8. Conclusions

Chit funds are traditional financial institutions that cater to the credit needs of different sections of the society in India. In our previous study, we have found that 5–10 per cent of the households in the urban area of the five states of Tamil Nadu, Karnataka, Andhra Pradesh, Kerala, and Delhi participate in registered chit funds. The average money circulated via chits is

around Rs. 50,000¹⁸. Chit funds can also be considered to be a tool for financial integration of the poor. It is a unique instrument that caters to the needs of both savers and borrowers. This industry, therefore, plays an important role in the Indian economy. However, there have been innumerable instances where this industry has been highly criticized for being unsafe and defaulting on its commitment to its members. This is primarily because this industry, unlike banks, is not covered by deposit insurance and does not have proper recourse when members default¹⁹. In light of the above, the onus of ensuring that the credit risk is minimized lies with the companies. A well-developed credit scoring mechanism would go a long way in mitigating such credit risk.

As we have seen, chit funds still majorly rely on human judgement while assessing the riskiness of their members. An analysis of the data collected from the two companies shows that 35 per cent of members have defaulted at least once during their tenure with the company, 24 per cent of whom have defaulted after taking a loan. This statistic highlights the inefficiency of the current risk assessment mechanism. We also find that the members participating in high value, long duration schemes are riskier indicating that the Value at Risk (VaR) could be quite significant.

An important first step in the direction of developing a better risk mitigation mechanism would be to collect comprehensive demographic and background details from the members at the time of admission and to also constantly update this data. Currently, we find that very little demographic data is being collected by most of the companies. Moreover, only around 30 per cent of the data is digitized, which means that the rest of the data is difficult to access and is highly unorganized. There is also a pressing need for standardization of collateral and surety requirements. There seems to be a high variation in the number of sureties collected from members, ranging from one to six. We also find that the most used collateral is that of other chits that the members participate in, indicating that the members might be strategically using the chits to make profits.

Illiquidity appears to be one of the major drawbacks of participating in chit funds. An average member participates in four auctions before winning the bid. This means that any urgent demand for credit is unmet by chits, thereby forcing the member to seek outside finance at possibly greater cost. This in turn increases the financial burden of the member which increases the probability of the member defaulting in the chit scheme.

From our analysis, we find that the longer a member stays with the company or the more number of schemes a member par-

¹⁸S. Buteau and P. Rao, *Chit Funds as an Innovative Access to Finance for Low-income Households*, Small Enterprise Finance Centre, Institute for Financial Management and Research, August 2010.

¹⁹Chit funds can lodge a case against defaulting members to recover the amount due. However, this procedure is time consuming and the probability of settlement is very low and delayed.

ticipates in has an increasing effect on default. Middle-aged and male members also show greater tendency to default. We find that, at the time of joining, the least risky types of members are government servants with low income, and self-employed professionals with high income. The groups with maximum riskiness are business owners and retired individuals with medium and high income. Borrowing at the early stages of a scheme, making large number of late payments or lump-sum payments, and participating in auctions frequently are also indicators of riskiness of a member. At the time of borrowing, the least risky group appears to be self-employed professionals with high income, and the riskiest group appears to be business owners with medium and high income.

The findings of this paper are primarily based on retrospective data that is currently available with the chit fund companies. The effectiveness of the model could be assessed by having the companies apply the scores on their prospective clients and borrowers, and study the default rates and delayed payments thereafter. There is considerable scope to fine-tune the model based on additional data that could be collected by the company at the time of admitting a member.

As we have discussed in the scalability section, there is an imminent need for the industry to move towards computerization of their data. Once their existing data has been digitized and they have a system in place to collect new information electronically, the credit scoring model could also be automated. At a much later stage, when the entire chit fund industry has adopted and automated the credit scoring model, a chit fund credit bureau could also be set up, whereby the scores of different chit fund members could be shared amongst the companies. Such scores could also be 'sold' to third parties like banks and other financial institutions. Therefore, a move towards a proper credit risk assessment mechanism would benefit not only the individual chit fund companies and the industry, but also the members and the community as a whole.

Supplementary material

Supplementary File 1: Tables for reference.

[Click here to access the data.](#)

Supplementary File 2: Glossary of terms.

[Click here to access the data.](#)

Data availability

Credit scoring data is available on Harvard Dataverse: <https://doi.org/10.7910/DVN/GWOTGE>²⁰

Data are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Competing interests

No competing interests were disclosed.

Grant information

Bill and Melinda Gates Foundation [OPPGD489].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

This research work would not have been feasible without the help of many. We are immeasurably grateful to Mr Sivaramakrishnan, the General Secretary of the All Indian Association of Chit Funds (AIACF) for his continuous support and outstanding dedication. Our boundless gratitude to the participating chit companies as well: The Balussery Benefit Chit Fund, Sri Nithyakalyani Chits, Shriram Chits, Kapil Chits, and Sri Rajeshwari Chits. We are also thankful to all the members companies of the AIACF for their support.

We owe much to our colleagues, Aneesh Mannava, Kothandam Devan, Anwar Basha, Ramakrishna Reddi, Satish Bidarkar, and various members from the data entry team who provided us with invaluable help at various stages of the research.

²⁰Rao, Preethi, 2018, "Credit Scoring data", <https://doi.org/10.7910/DVN/GWOTGE>, Harvard Dataverse, V1, UNF:6:WvEu6TmqM5NaSc99s6WnyQ==

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 11 June 2018

<https://doi.org/10.21956/gatesopenres.13827.r26426>

© 2018 Klonner S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stefan Klonner 

South Asia Institute, Department of Development Economics, Heidelberg University, Heidelberg, Germany

This is an interesting, innovative paper. The research questions is: what are the predictors of borrowers' riskiness in chit funds, a popular semi-formal savings and credit group scheme in India, where a non-banking financial company effectively acts as the lender. Riskiness refers to the probability of default of a borrower. The paper also has a practical component as it presents credit scores, which – as the authors suggest – chit fund companies could use to decide on the admission of certain types of individuals into chit funds and the degree of screening, collateralization and monitoring once a loan is awarded to a chit fund member. The data is very original and comes from chit fund companies. In addition to data available in data bases of these companies, the authors have entered information that these financial companies typically keep only in paper form.

The research question is important and innovative, assessing credit risk with statistical methods outside the domain of bank lending in a developing country. The paper is overall well-written and accessible. However, I have a major methodological concern, which is relevant for the authors' main conclusions:

As the authors state, an organizer mixes customers with different needs, savers and borrowers in a group. Different occupational groups are savers and borrowers to systematically different extents, it seems. On the other hand, default and arrear mechanically differ across rounds in the same chit fund and also across chit value-duration combinations. If e.g. self-employed participants tend to lift earlier pots and earlier pots are mechanically more prone to default because there is a greater debt amount and longer repayment spell, then a finding that self-employed are more risky in the authors' regression framework could be driven merely by their taking of earlier pots. In contrast, in my understanding, a credit score ranks borrowers' propensity to default conditional on the SAME loan product. To account for this the authors should explore the use of round and value/duration-combination fixed effects in their econometric analysis. Then all results would be conditional on the specific loan product, where a loan product is a given round in a specific value-duration combination.

Another concern regarding the writing is that the paper is sometimes not written like a focused research paper but more like a report, where a lot of facts are enumerated without being put into a specific analysis framework and flow of argument.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Development economics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 29 May 2018

<https://doi.org/10.21956/gatesopenres.13827.r26424>

© 2018 Zhang Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Yu Zhang

Department of Accounting and Finance, Eastern Michigan University, Ypsilanti, MI, USA

This is an interesting article on credit analysis of chit funds in India. This research has collected a large dataset to summarize the demographics and operations of chit funds in India. The whole article has a clear layout with logical analysis. It adds contribution to the chit funds literature with first-hand data collection and analysis.

There are some suggestions listed below:

1. There is no literature review. Unless this is the only paper that studies chit funds, there should be some sort of literature review in this area. Without reading the literature review it is hard to specify the originality level of the paper and its contribution to the literature.
2. Introduction can be more direct: it takes me to the end of the introduction to finally know the main purpose of the paper.
3. In the OLS regressions, the paper created a lot of dummy variables based on occupation and income level together, such as business owner, low income; Business owner, medium income, etc. Another way to do it is to create dummy variables on occupations and categorical variable on income level with numbering from 1 to 3; and then create interaction terms which are the multiplication of the two. By doing that, we can see whether certain occupation by itself (without considering income) will have a significant effect on default rate. Same thing applies to income level by itself. The interaction terms achieve the same purpose as the paper did. This is just a suggestion and is not a necessity.
4. In the OLS regression, please report the F test of overall significance. Only R2 is reported and we cannot know whether the models have a good fitness or are reliable.
5. A minor thing is in table 16, column 4, is that % of unsuccessful bids? It is now written as % age of unsuccessful bids, which is quite confusing.

To sum up, it is an interesting paper and I recommend to accept it with minor modification of adding literature review if there is any, and other changes.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.