Essays on Credit Access and Household Finance

by

Seyma Kalay

Matricola 1050336

A thesis submitted in partial fulfillment

for the degree of

Doctor of Philosophy

Department of Economics

Bergamo University

Graduate School of Economics and Quantitative Methods, Bergamo

Supervisors:

Prof. Michela Cameletti

Prof. Federica Maria Origo

2020

Declaration

I, Seyma Kalay, hereby declare that this Ph.D. thesis titled " Essays on Credit Access and Household Finance" is a presentation of my original research work for the degree of Doctor of Philosophy in Economics under the guidance and supervision of Prof. Michela Cameletti and Prof. Federica Maria Origo.

I affirm that this work has not been submitted previously in the same or similar form to another examination committee at another department, university, or country.

Signed: _____

Date: _____

By three methods we may learn wisdom: First, by reflection, which is noblest; Second, by imitation, which is easiest; and third by experience, which is the bitterest.

CONFUCIUS (551BC − 479BC)

Acknowledgments

Abstract

Determinants of access to credit/finance[1] have been long discussed in the literature. Previous studies have shown the importance of access to finance and how limited access to finance has a negative impact on economic growth. This thesis contributes to the literature by exploring the literature on access to credit. It contributes characteristics of access to credit, suggests policy implementations, and provides interactive visualization interfaces.

The contributions are discussed in three self-contained chapters:

Chapter 2 reviews the literature on determinants of access to finance, using a meta-literature review known as quali-quantitative (e.g. content-bibliometric) analysis. Through the bibliometric and content analysis, the chapter presents the influential features of the literature under determinants of finance such as countries, affiliations, journals, authors, articles, trend papers, and determines two main research streams, and provides future research questions.

Chapter 3 studies the determinants of access to credit of Chinese households, for formal, informal, and both sources by implementing machine learning (ML) techniques into different data-splits. Using predictive modeling, rather than explanatory modeling, helps us to get more robust results. The chapter finds CCP.0 & CCP.1 data split achieves a higher predictive power, and most importantly the other data-splits as well the benchmark data-set can be misleading to characterize the Chinese households.

Chapter 4 clusters Chinese households based on their financial strength, finds access to credit is already towards to better of, and introduces an interactive map to identify advantage and disadvantage households for the future policy implementations.

---

[1]Note: Throughout this study credit and finance are used interchangeably.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

Access to finance, often defined as access to financial services (e.g., formal account and saving), meets the needs of individuals and businesses at an affordable cost. The usage of financial services, which benefits individuals, SMEs, and governments, has been associated with the stability of the economy and well-being. In the loan market, however, discrimination occurs based on the participant characteristics (such as gender, ethnicity, political affiliation, income, etc.) even after controlling relevant factors (i.e., the credit history of the participant). For a constant increase in economic growth, the government's aim should be toward economically disadvantaged groups to promoting them.

The contributions of each chapter as follows;

In Chapter 2, we present a state-of-the-art review on scientific production on "determinants of finance", using bibliometric techniques coupled with content analysis. We highlight the influential features of the literature under determinants of finance, such as countries, affiliations, journals, authors, articles, and trend papers. We also present co-citation network and we determine two main research streams (i) lending to small borrowers (ii) lending to big borrowers. Through the bibliometric and content analysis, we provide 13 future research questions. Additionally, we created a link for the bibliometric analysis where the results can be reproducible in Shinyapp[1].

In Chapter 3, we examine access to finance for formal, informal, and both sources for households' heads for different data-splits to define the best data-split that has a higher predictive power among others. Following James et al. (2013), we used 10 cross-validation (cv) based on 80/20 train/test set and we run, generalized & multinominal logistic models and 3-ML models namely; bagging, random forest, and gradient boosting for both $y = Access\,Loan$ & $y = Loan\,Type$ using 3 financial asset variables. The findings suggest that access to formal loan is low and CCP.0 & CCP.1 data split achieves the highest predictive power to explain accessing the loan and its type.

---

[1]See at: https://seymakalay87.shinyapps.io/biblio/3

In Chapter 4, firstly, we implement unsupervised learning (K-means clustering) to distinguish wealthy and poor households. Secondly, we implement supervised learning (e.g., bagging, random forest, and gradient boosting) and multinominal logistic model to predict and to confirm the goodness of the cluster using 10 cv with based on 80/20 train/test set. Our findings suggest that access to formal loan is low and it is associated with the households' financial strength. Using user interface[2], we map advantaged and disadvantaged households to help policy maker to expand financial inclusion especially among socially and economically disadvantaged groups.

---

[2]See at: https://seymakalay87.shinyapps.io/micro/

Chapter 2

Determinants of Access to Finance: A Bibliometric Literature Review

2.1   Introduction

In academic research, the literature review plays a crucial role in gathering accumulated body of knowledge and guiding the future research directions (Cropanzano, 2009; Kunisch et al., 2018), regardless of discipline. But, the number of research papers is gradually increasing and it is becoming infeasible to remain up-to-date. Typically, prior to conducting a new research, researchers review the literature to state the available evidence on a topic. Often a literature review, the selection of some studies over others, is not fully representative of ever-growing knowledge but an arbitrary selection, which leads to choose a non-random data sample for further analysis. Consequently, narrative reviews often provide and offer a non-comprehensive background for testing and theory development.

The building block of academic research is to construct a research on and relate it existing studies. There already exist different types of literature reviews, such as integrative (Baumeister and Leary, 1997; Wong et al., 2013), systematic, and meta-analysis (Davis et al., 2014; Liberati et al., 2009; Moher et al., 2010). Depending on the purpose of the researcher (shown in Table 2.1), all types of literature reviews can be helpful (Snyder, 2019). To synthesize and consolidate the past findings effectively, this paper will conduct meta (quali-quantitative) analysis which is defined as "the analysis of analysis" by Glass (1976), is a combination of quali-quantitative (i.e., content-bibliometric) analysis.

Bibliometric analysis helps to understand the scientific production, intellectual networks, trends, and publication patterns between the scholars, institutions, and countries since it is based on statistical measurements (Bourdieu and Farage, 1994; Broadus, 1987; Liu et al., 2014; Pinto et al., 2014; Pritchard et al., 1969). Bibliometric analysis conducts reproducible literature review concept

(Broadus, 1987; Diodato and Gellatly, 2013; Pritchard et al., 1969) with more accurate and more objective analysis (Aria and Cuccurullo, 2017). Yet bibliometric methods are not new (Kessler, 1963; Small, 1973), they recently started to grasp attention from the databases (e.g., Thomson Reuters Web of Science (Wos), Scorpus, etc.) and software developers. Although there are a number of statistical tools to conduct a bibliometric study (such as CitNetExplorer, SciMAT, VOSviewer, BibExcel, Sci2, CiteSpace, etc.) only R can provide a complete bibliometric workflow (Aria and Cuccurullo, 2017).

| | Approaches | | |
	Systematic	Semi-systematic (Meta-analysis)	Integrate
(1) Purpose	Synthesize and compare evidence	Overview research area and track development over time	Critique and synthesize
(2) Research question	Specific	Broad	Narrow or broad
(3) Research strategy	Systematic	May or may not be systematic	Usually not systematic
(4) Characteristics	Quantitative articles	Research articles	Research articles, books, and other published documents
(5) Analysis	Quantitative	Qualitative/Quantitative	Qualitative
(6) Contribution	Evidence of effect, Inform policy and practice	State of knowledge, Themes in literature, Historical overview, Research agenda, Theoretical model	Taxonomy or classification, Theoretical model or framework

Source at: Literature review as a research methodology: An overview and guidelines.

Table 2.1: Literature review approaches.

There are a wide range of bibliometric analyses[1] from many multidisciplinary fields such as international economics (Teixeira and Carvalho, 2014), risk management (de Araújo Lima et al., 2020), financial econometrics (Baker et al., 2020), bankruptcy prediction (Shi and Li, 2019), and so on. As per best of our knowledge, no such research has previously combined bibliometric and content analysis under the "determinants of finance".

Therefore, the aim of this study is to explore meta data-set of all the papers indexed in the Web of Science (WoS) under the "determinants of finance"by applying the complete bibliometric workflow (Zupic and Čater, 2015) coupled with content analysis. Additionally, using R open-

---

[1]See the list of the publications, using bibliometric methodology in R: https://www.bibliometrix.org

source, we provide the code and meta data-set for this study[2] as well as a reference for Shinyapp[3] to demonstrate a reproducible bibliometric analysis.

This study answers three main questions: (1) What are the influential features of the "determinants of finance" literature? WWe identify the top ten influential countries (Table 2.2), affiliations (Table 2.3), journals (Table 2.4), authors (Table 2.5), articles (Table 2.6), and trend papers (Figure 2.2) since scientific products are heavily dependent on the collaboration between scholars (Acedo et al., 2006; Zhai et al., 2014), affiliations and countries (Finardi and Buratti, 2016). (2) What are the main research stream of "determinants of finance"? We analyze the co-citation (Figure 2.3), co-authorship (Figure 2.4), and co-word (Figure 2.5 and Table 2.8) coupled with content analysis and identify two main research streams (Table 2.8) (i) lending to small borrowers and (ii) lending to big borrowers. (3) What are the future research questions under this topic? With bibliometric and content analysis, we identify 13 future research questions (Table 2.7).

The rest of the chapter is structured as follows. Section 2.2 describes the methodology. Section 2.3 presents the data collection. Section 2.4 shows the results of the analysis. Section 2.5 shows the data visualization. Section 2.7 introduces the Biblio Shinyapp. Section 2.8 presents the limitations of the bibliometric study, and Section 2.9 presents the literature review of research streams.

2.2 Methodology

"If I have seen further, it is by standing upon the shoulders of giants."

- Isaac Newton

The importance of bibliometric analysis goes beyond the conventional literature review and a huge number of publications makes crucial to utilize bibliometric analysis in order to observe the trends, emerging topics, research streams, and leading scholars over a time span. Bibliometric analysis helps researchers to understand the emergent and evolutionary trends over the years (Merediz-Sola and Bariviera, 2019).

The quali-quantitative literature review, consisted of multiple steps and sample selection which will be discussed in Section 2.3. Additionally, Figure 2.1 provides the complete snapshot of the

---

[2]See at: https://github.com/seymakalay, under biblio repository.
[3]See at: https://seymakalay87.shinyapps.io/biblio/

methodology as well as 5-stage bibliometric workflow[4].

2.2.1 Complete (5-stage) bibliometric workflow

A standard workflow to perform bibliometric analysis consists of five-stage (Zupic and Čater, 2015). **(i)**-*Study design:* After defining the research question(s), with the help of bibliometric analyses, researchers can (a) recognize the main grasp of the topic as well as its intellectual structure, (b) capture the conceptual structure of the topic, and (c) create a social network in the research community. **(ii)**-*Data collection:* scholars collect the meta-data from the databases such as WoS, Scorpus, Google Scholars, and Science Direct. **(iii)**-*Data analysis:* the descriptive analysis and the creation of bibliometric coupling network start (a) *co-citation analysis:* helps to identify the research stream (Kim and McMillan, 2008) and shows the similarity between the manuscripts, affiliations, journals, and authors (Aria and Cuccurullo, 2017), (b) *co-author (collaboration) analysis:* shows the collaboration network among the authors (Glänzel, 2001; Liu et al., 2005; Peters and Van Raan, 1991), (c) *co-word (co-occurrence) analysis:* can be applied to the keywords, manuscript's titles, abstracts, and topics which demonstrates the most important words for that research field (Callon et al., 1983). **(iv)**-*Data visualization:* visualizing network mapping and **(v)**-*Interpretation.*

---

[4]Complete (5 stage) bibliometric workflow is based on the idea of Zupic and Čater (2015).

Step 1

Objectives

of the Study

→

1) What are the influential features of the "determinant of finance" literature?

2) What are the main research stream of "determinants of finance"?

3) What are the future research questions under this topic?

-Complete (5 Stage) Bibliometric Workflow[4]-

(i)-*Study design:*

1) Deciding the research question. In or case "determinants of finance".

(ii)-*Data collection:*

2) Selecting a database that contains the bibliometric data (e.g., WoS).

3) Articles were selected based on the set of keyword search which are relevant to "determinants of finance" (e.g., discrimination, loan, financial inclusion etc.).

4) Filtering to search (e.g., English, aricles, SSCI, SCI .etc)[a].

5) The abstract of each article was read to decide if "determinants of finance" is not marginally mentioned but have a direct content in the article.

6) Finally, 210 articles were selected to conduct bibliometric analysis.

(iii)-*Data analysis:*

7) After exporting .txt file from WoS database, we converted it into a bibliographic dataframe using R and contacted descriptive statistics.

(iv)-*Data visualization:*

8) Bibliographic dataframe pave the way to network matrix then, network mapping can be conducted.

(v)-*Interpretation:*

9) Interpretation is the last stage of the complete bibliometric workflow which helps to researchers make sense of bibliometric's results.

---

[a]Hence Shinyapp[3] considers each uploaded articles unless bibliometric meta-data is missing. Additionally, filtiring for the data colection should be done in WoS database (in our case, published English papers with SSCI & SCI indexes).

Step 2

Bibliometric

Analysis

→

Step 3

Contributions

of the Study

→

1) We find the influential features of "determinants of finance" literature such as countries, affiliations, journals, authors, articles, and trend papers.

2) We identify two main reserch streams under "determinants of finance" literature namely, (i) Lending to small borrowers and (ii) Lending to big borrowers.

4) We print the co-citation, co-authorship, and co-word network analysis.

5) Using bibliometric and content analysis, we identify 13 future research questions.

6) We provide the code[2] and produce an Shinyapp[3] for the reproducibly.

Figure 2.1: Methodological approach provides a snapshot for this study. Source: author's presentation.

2.3 Data collection

This study adopts a three steps systematic sample selection to fulfill study objectives (see data collection in Step 2 at Figure 2.1). In order to run bibliometric analysis, the first step is to select the data-base to collect relevant articles. We select Web of Science Core Collection (WoS) data-base which is one of the most well known academic databases and includes more than 59 million records from various disciplines (e.g., economics, mathematics, engineering, and computer science) since 1950 and presents Science Citation Index Expanded (SCI-EXPANDED) and Social Sciences Citation Index (SSCI) since 1985.

The second step is filtering the literature search that covers the largest scoop of the literature. We search from WoS, using relevant set of keywords (e.g., financial inclusion, discrimination and access to finance etc.). At the first level, we applied filtering to the search; English published articles with SSCI and SCI-EXPANDED indexes.

At the third step is to include/disregard relevant/irrelevant articles which will affects the results of the bibliometric analysis. By reading all the abstracts of published English articles, we considered to include the relevant articles if only "determinants of finance" is not marginally mentioned, but has a direct content in the article. The articles that were use for the bibliometric analysis and the affiliation distribution for each journal can be found in Appendix A, sorted alphabetically.

In total, we considered 210 articles relevant to "determinants of finance" within the bibliometric meta-data, the majority of the disciples belong to economics (64%), business & finance (23%), development studies (17%) business (16%), management (10%), and other multidisciplinary areas.

2.4 Data analysis

Within the meta data-set, 210 articles are published by 116 different sources (e.g., journals, books, etc.) and only 37 of the articles are single-authored. The average number of authors per article is 2.15, with 466 total authors. Average citations per article is about 61.59, and TGC, for all the 210 papers, is 13362 during the period 1993-2021.

Descriptive statistics help us to identify the most influential features of "determinants of finance" such as countries (Table 2.2), affiliations (Table 2.3), journals (Table 2.4), authors (Table 2.5), articles (Table 2.6), and trend papers (Table 2.2). Additionally, the keywords to search the literature

on "determinants to finance" are confirmed through co-word analysis (Table 2.8 and Figure 2.5). We present future research questions in Table 2.7 and the articles, in alphabetical order, that we used for the bibliometric analysis in Appendix A.

Also, it is important to mention the key terms which will be used for bibliometric analysis. **TLC:** total local citation which presents the total number of citation by the authors which are in our meta-data (in this case 210), **TLC/t:** total local citation per year, **TGC:** total number of global citation, **TGC/t:** total global citation per year, **SCP:** single country, **MCP:** multiple country publications, **Year:** publication year, and **Article.No.:** the total number of published English articles.

2.4.1 Most influential countries and affiliations

The concept of determinants of finance is a long-discussed phenomenon. Therefore, it is important to determine the center of excellence in terms of countries and affiliations. We present in Table 2.2 and Table 2.3 the top 10 countries and affiliations, sorted by Article.No. and TLC, respectively, with a few detailed information.

	Country	Article.No.	%Freq	SCP	%SCP	MCP	%MCP	TGC	TGC/t
1	USA	83	41.29	64	47.76	19	28.36	8120	97.80
2	United Kingdom	26	12.93	19	14.18	7	10.45	1024	39.40
3	China	14	6.97	7	5.22	7	10.45	421	30.10
4	Italy	7	3.48	5	3.73	2	2.99	125	17.90
5	India	6	2.99	5	3.73	1	1.49	354	59.00
6	France	5	2.49	2	1.49	3	4.48	206	41.20
7	Germany	5	2.49	2	1.49	3	4.48	491	98.20
8	Netherlands	4	1.99	2	1.49	2	2.99	502	125.50
9	New Zealand	4	1.99	4	2.99	0	0.00	142	35.50
10	South Africa	4	1.99	4	2.99	0	0.00	32	8.00

Note: The table is sorted based on total number of Article.No. %Freq, %SCP, and %MCP are the percentage of the total Article.No., SCP, and MCP, respectively.

Table 2.2: The most influential countries

Table 2.2 shows the countries sorted by Article.No. The highest total number of publications (Article.No.: 83) and total global citations (TGC: 8120) is in the USA which is equivalent to 41.29% of the overall publications in our meta data-set, with an average of about 97.80 citations per year, and 64 of those (47.76%) with single country publications (SCP) and 19 of those (28.36%) with multiple country publications (MCP).

Netherlands has 4 articles and a moderate TGC: 502 comparing to top three countries, but it has the highest TGC/t: 125.50 which indicates in Netherlands articles have gained popularity by the scholars. Whereas the number of articles seems a lot for the top 3 countries (USA, United Kingdom, and China) their TGC/t and TGC per article, however, is less compared to Netherlands.

	Affiliation	Articles	TLC	TLC/t	TGC	TGC/t
1	World Bank	14	69	6.24	2558	197.35
2	Georgetown Univ	2	53	2.55	297	14.62
3	Dartmouth Coll	1	31	1.72	237	13.17
4	Wellesley Coll	1	31	1.72	237	13.17
5	Tilburg Univ	5	30	3.13	634	55.68
6	Fed Reserve Syst	2	28	1.48	239	12.91
7	Harvard Univ	5	28	2.58	1963	144.34
8	Ctr Naval Anal	1	26	1.13	111	4.83
9	German Inst Econ Res Diw Berlin	2	22	2.34	195	28.75
10	Robert Gordon Univ	2	22	1.89	177	15.64

Note: The table is sorted based on TLC. Appendix A.2 shows the universities published in these affiliations.

Table 2.3: The most influential affiliations

The ranking of affiliations is presented in Table 2.3. On the one hand, the top 3 affiliations are World Bank (TLC: 69), Georgetown Univ (TLC: 53), and Dartmouth College (TLC: 31) which are all in the USA. On the other hand, Harward Univ (TLC: 28) has been more appreciated by the researchers with TGC/t: 144.34 for 5 articles.

The top 10 countries (Table 2.2) and affiliations (Table 2.3) can be considered as "centers of excellence" for previous studies on determinants of finance. These findings may be useful for those who are interested to collaborate on research projects and to organize workshops on determinants of finance.

2.4.2 Most influential journals and authors

Identifying the top journals and authors will be useful for the scholars working on the determinants of finance to create collaboration between authors and to target the top journals in the field. Table 2.4 and Table 2.5 show the top 10 influential journals and authors, both sorted based on three criteria: (1) total number of published articles (Article.No.), (2) total local citation received per year (TLC/t), and (3) total global citation received per year (TGC/t).

The results, in Table 2.4, indicate that World Development journal has the highest Article.No. (14) and TLC/t (6.49). Whereas, Management Science, Quarterly Journal of Economics, and

	Journal	Article.No.	TLC	TLC/t	TGC		TGC/t
1	World Development	14	50	6.49	683		88.75
2	Journal Of Development Studies	10	18	2.29	200		23.32
3	Journal Of Banking & Finance	8	23	2.28	1220		96.72
4	Small Business Economics	7	18	2.39	230		29.86
5	Environment And Planning A	5	1	0.07	194		11.72
6	Journal Of International Development	5	5		188		
7	Sustainability	5	4	2.50	46		25.83
8	Emerging Markets Finance And Trade	4	6	1.29	68		10.49
9	Entrepreneurship Theory And Practice	4	15	1.15	729		56.38
10	Finance Research Letters	4	2	0.53	76		29.07
	Journal	Article.No.	TLC	TLC/t	TGC		TGC/t
1	World Development	14	50	6.49	683		88.75
2	Journal Of Financial Intermediation	1	13	2.60	118		23.60
3	Journal Of Business	3	44	2.51	479		28.86
4	Sustainability	5	4	2.50	46		25.83
5	Small Business Economics	7	18	2.39	230		29.86
6	Journal Of Development Studies	10	18	2.29	200		23.32
7	Journal Of Banking & Finance	8	23	2.28	1220		96.72
8	China Economic Review	2	12	2.00	92		15.34
9	Journal Of Comparative Economics	2	24	1.96	299		23.08
10	American Economic Review	2	16	1.90	81		9.61
	Journal	Article.No.	TLC	TLC/t	TGC		TGC/t
1	Journal Of Banking & Finance	8	23	2.28	1220		96.72
2	World Development	14	50	6.49	683		88.75
3	Journal Of Financial Economics	2	16	1.23	917		70.54
4	Management Science	1	2	0.29	439		62.71
5	Entrepreneurship Theory And Practice	4	15	1.15	729		56.38
6	Quarterly Journal Of Economics	1	14	0.88	866		54.12
7	Annual Review Of Sociology	1	4	0.31	689		53.00
8	Journal Of Finance	2	5	0.22	812		52.86
9	Small Business Economics	7	18	2.39	230		29.86
10	Finance Research Letters	4	2	0.53	76		29.07

Note: The table is sorted by Article.No. (top), TLC/t (middle), and TGC/t (bottom).

Table 2.4: The most influential journals

Annual Review of Sociology seems to be more appreciated by the researchers, having TGC/t 62.7, 54.12, and 53 per article, respectively. Additionally, in Appendix A.2 we can see the authors' universities who published in each journal.

Table 2.5 shows that in terms of productivity, the top three first-authors with the maximum number of published articles (Article.No.: 5, 3, and 2) are Beck T., Wyly Ek., and Agier I. On the one hand, the 5 articles of Beck T had received 1111 total global citations (TGC) with an average yearly TGC = 83.23. On the other hand, TGC/t of Fernandes D., Khwaja Ai., Pager D., and Campbell Jy. are 62.71, 54.12, 53, and 51.27, respectively, for only 1 article.

	1st Author	Affiliation	Article.No.	TLC	TLC/t	TGC	TGC/t
1	Beck T	World Bank	5	21	1.79	1111	83.23
2	Wyly Ek	Rutgers State Univ	3	4	0.23	148	8.97
3	Agier I	Univ Libre Bruxelles	2	7	0.87	96	12.00
4	Allen F	Imperial Coll London	2	14	2.74	175	31.74
5	Asiedu E	Univ Kansas	2	16	1.90	81	9.61
6	Bates T	Wayne State Univ	2	1	0.04	38	3.80
7	Bayer P	Duke Univ	2	0	0.00	25	7.58
8	Black Ha	Univ Tennessee	2	6	0.33	26	1.43
9	Carter S	Univ Sterling	2	9	0.83	283	29.26
10	Cavalluzzo Ks	Georgetown Univ	2	53	2.55	297	14.62

	1st Author	Affiliation	Article.No.	TLC	TLC/t	TGC	TGC/t
1	Allen F	Imperial Coll London	2	14	2.74	175	31.74
2	Cavalluzzo Ks	Georgetown Univ	2	53	2.55	297	14.62
3	Aterido R	World Bank	1	16	2.00	83	10.38
4	Asiedu E	Univ Kansas	2	16	1.90	81	9.61
5	Beck T	World Bank	5	21	1.79	1111	83.23
6	Blanchflower Dg	Dartmouth Coll	1	31	1.72	237	13.17
7	Muravyev A	Inst Study Labor	1	20	1.67	145	12.08
8	Fungacova Z	Bank Finland Inst Econ Transit Bofit	1	9	1.50	64	10.67
9	Zins A	Univ Strasbourg	1	7	1.40	89	17.80
10	Alesina Af	Harvard Univ	1	11	1.38	81	10.12

	1st Author	Affiliation	Article.No.	TLC	TLC/t	TGC	TGC/t
1	Beck T	World Bank	5	21	1.79	1111	83.23
2	Fernandes D	Erasmus Univ	1	2	0.29	439	62.71
3	Khwaja Ai	Harvard Univ	1	14	0.88	866	54.12
4	Claessens S	Int Monetary Fund	2	16	1.21	711	53.85
5	Pager D	Princeton Univ	1	4	0.31	689	53.00
6	Campbell Jy	Harvard Univ	1	1	0.07	769	51.27
7	Allen F	Imperial Coll London	2	14	2.74	175	31.74
8	Carter S	Univ Sterling	2	9	0.83	283	29.26
9	Hastings Js	Brown Univ	1	2	0.25	198	24.75
10	Houston Jf	Univ Florida	1	1	0.14	172	24.57

Note: The table is sorted by sorted based on Article.No. (top), TLC/t (middle) and TGC/t (bottom).

Table 2.5: The most influential authors

2.4.3 Most influential articles/topics

Research direction (in any field of study) is determined by the influential articles/topics which might be helpful to understand the scope of the research streams. Table 2.6 shows the 10 most influential articles sorted by TLC/t and TGC/t, respectively.

Additionally, bibliometric analysis coupled with content analysis helps us to identify several research directions. Table 2.7 shows us a few key research questions which were retried from the top influential papers: (1) What type of approaches to financial education are required? (2) What kind of policies should be implemented to ease the women access to finance? (3) How the stock return or debt reduction differs for the firms, having political ties? (4) How GDP and welfare of a

country are affected negatively, when the companies are favor in access to finance?

In addition, with the help of content analysis, we define nine more research questions, (5) Is it possible to have more fare and equal financial system? (6) What kind of policies must target minor groups? (7) Can house owning or land titling programs increase investment on education and helps to access to finance in the long run? (8) How human capital is affecting approval of loan? (9) Comparing banking-relationship in develop and developing countries for lending decision making. (10) Why banking-relationship is more important for developing countries? (11) What is the financial accessibility today? (12) What is the role of digital finance? (13) How access to finance differs between countries where the higher education is free?
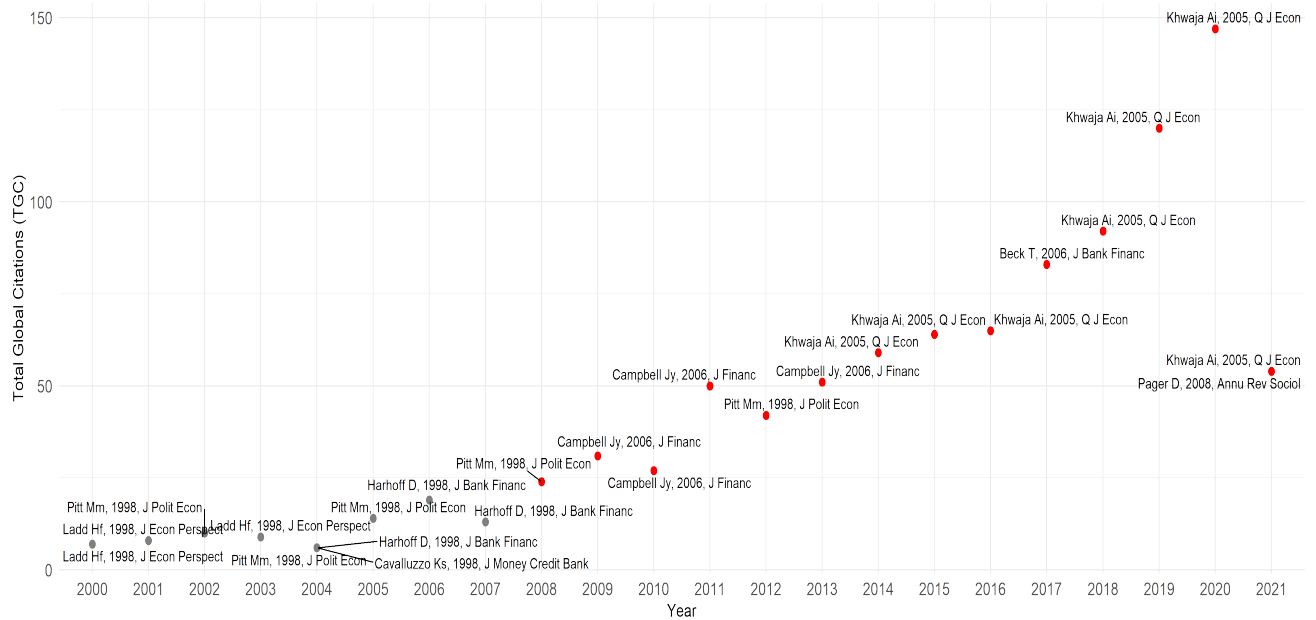
Figure 2.2: Trend Papers, total global citation (TGC on the y-axsis) exceeds yearly 20 (TGC > 20) is shown in red, otherwise gray. Only the last 20-year (on the x-axsis) is plotted to make results readable.

Figure 2.2 plots the articles based on the highest total global citation (TGC) received based on the year shown on the x-axis, which suggests that researchers may consider these articles for designing future studies. The total global citation (TGC) for the articles has an increasing trend in time and article Khwaja Ai. (2005) has gained the attention of the scholars, received its highest TGC in early 2021 which match with the Table 2.5.

2.5   Data visualization

In this section, the determinants of finance literature are visualized through bibliometric analysis by applying co-citation, co-word, and co-authorship network analysis.

2.6   Bibliometric coupling

Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.) such that the connection of different attributes can be presented through the document-attribute matrix denoted by X, where each row presents a document (D) and each column presents an attribute (A). The generic element of matrix X is $x_{ij} = 1$ if the i-th document (D_i) has the j-th attribute (A_j), otherwise $x_{ij} = 0$.

Let n is the total number of documents (rows) and $\sum x_{+j}$ is the sum of the total number of documents which has the j-th attribute shown as

$$\sum x_{+j} = \sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} D_i A_j = D_1 A_j + D_2 A_j + ... + D_n A_j$$

where m is the total number of attributes (columns) and $\sum x_{i+}$ is the sum of the total number of attributes which is in the i-th document shown as

$$\sum x_{i+} = \sum_{j=1}^{m} x_{ij} = \sum_{j=1}^{m} D_i A_j = D_1 A_1 + D_1 A_2 + ... + D_1 A_m$$

According to Kessler (1963), two articles are called bibliographically coupled when there is at least one commonly cited source in the reference lists of both documents, and this relationship gets stronger when the common cited sources increase.

The general formula of bibliometric coupling matrix can be written as

$$B_{coup} = X X^T$$

2.6.1   Co-citation network

The purpose of co-citation network is to analyze the commonalities, connection, strength of the articles, authors, or journals. The logic behind co-citation analysis is that the content of two articles

are similar when they are cited together and co-citation matrix can be obtained by

$$B_{cocit} = X^T X$$

where X is a *Document x Cited reference* matrix and element b_{ij} shows the number of total co-citation between documents i and j.
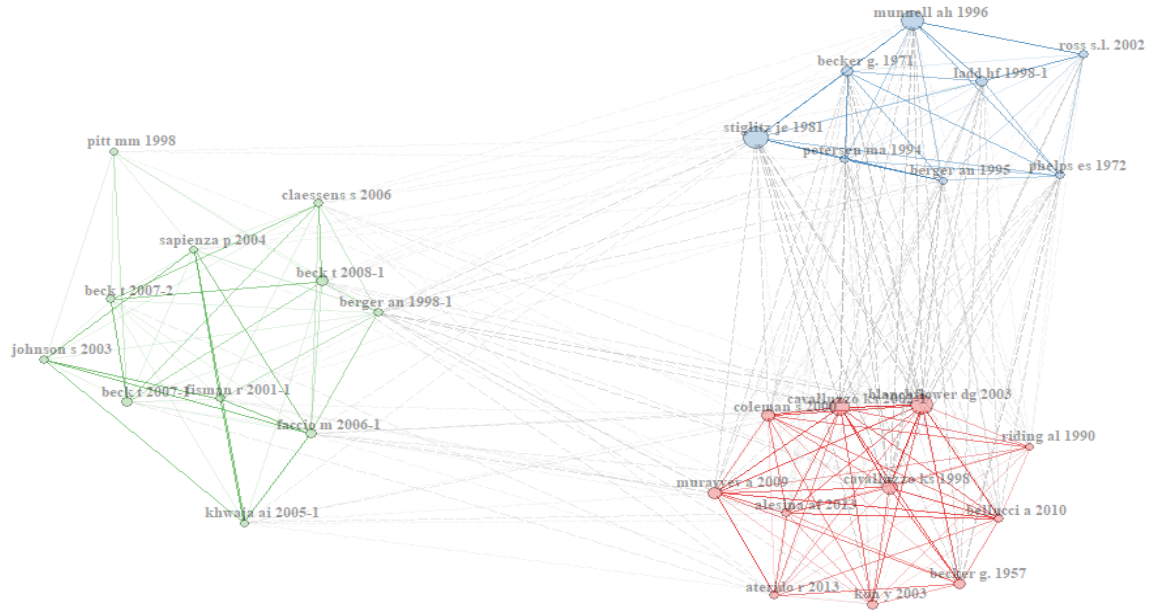
Figure 2.3: Stream/Cluster visualization in "determinants of finance" though bibliometric co-citation analysis between 1950 and 2020. (i) Lending to small borrowers (in blue) and (ii) Lending to big borrowers (in red).

Figure 2.3 presents the co-citation network of articles in the form of multiple clusters. The same color articles are said to be in the same research stream or cluster. Moreover, the magnitude of the node is positively associated with the number of citations. The higher the number of citations the bigger the node. The result of co-citation shows that the determinants of finance literature are divided into three different streams. Discrimination in mortgage market, small businesses, and firms are in blue, red, and green clusters, respectively. For the simplicity, we will consider as two main streams: (i) lending to small borrowers (in blue and red) and (ii) lending to big borrowers (in

green).

2.6.2 Co-authorship network

The purpose of the co-authorship network is two-fold: (a) to identify the authors who are working on the same field of interest to planned future projects, (b) to recognize the authors who gave good contributions to the field. Co-authorship network can be obtained by

$$B_{coauth} = X^T X$$

where X is a *Document x Author* matrix and element b_{ij} shows the number of total collaboration exist between authors (e.g., researchers, research institutions, and countries) i and j.

Figure 2.4: The relationship between authors in "determinants of finance" though bibliometric co-authorship (collaboration) analysis between 1950 and 2020.

Figure 2.4 demonstrates a strong social network among the most productive authors who are working on "determinants of finance". Authors within the same cluster have at least one published

English articles and authors' collaboration can be seen by the links whereas node size indicates higher productivity. We can see the network connection between authors if the author has at least one article with his/her colleague.

Based on Figure 2.4 what we can say that Bect T. and Cull R. have a higher collaboration network between the authors. Additionally, we can say that Demirguc-Kunt A. has a direct relationship between Beck T., Klapper I., and Allen f and Cull R. and Bect T. seems to have a bigger quantity of articles comparing the other authors in the network cluster.

2.6.3 Co-word network

Co-word network shows the most used words in keywords, manuscript's titles, abstracts, and topics[5] for each stream. Exploring the research stream (Table 2.8) is important. Identifying keywords allows describing the content in the selected articles. Co-word network can be obtained by

$$B_{coword} = X^T X$$

where X is a *Document x Word* matrix and element b_{ij} shows the total number of co-occurrences between words (e.g., keywords, manuscript's titles, abstracts, and topics) i and j.

Figure 2.5 presents the most used keywords over the period of 1993 and 2020 through the co-word network analysis and Table 2.8 prints out these keywords for each research stream ((i) in blue and red (ii) in green), founded using co-citation network (Table 2.3).

2.7 Biblio

Biblio is a Shinyapp[3] which provides an interface for bibliometric analysis where the results of this study can be reproducible[6]. As mentioned in Section 2.2.1, to be able to conduct bibliometric analysis in Shinyapp[3], one needs **(i)-*Study design*** and **(ii)-*Data collection***.

Shinyapp is self-explained in the provided link. The app default dataset is a .txt file where it can be retrieved from WoS database for Scorpus database users select "Load bibliometrix file(s)" and upload the .bib file. Users should wait until "Upload Complete" then click Start Convention.

---

[5]We implement co-word analysis only in keywords, readers who are interested can find all the options in Shinyapp[3].
[6]See at: https://github.com/seymakalay/biblio/blob/master/biblio.2.R to retrieve xtables in LaTeXformat.
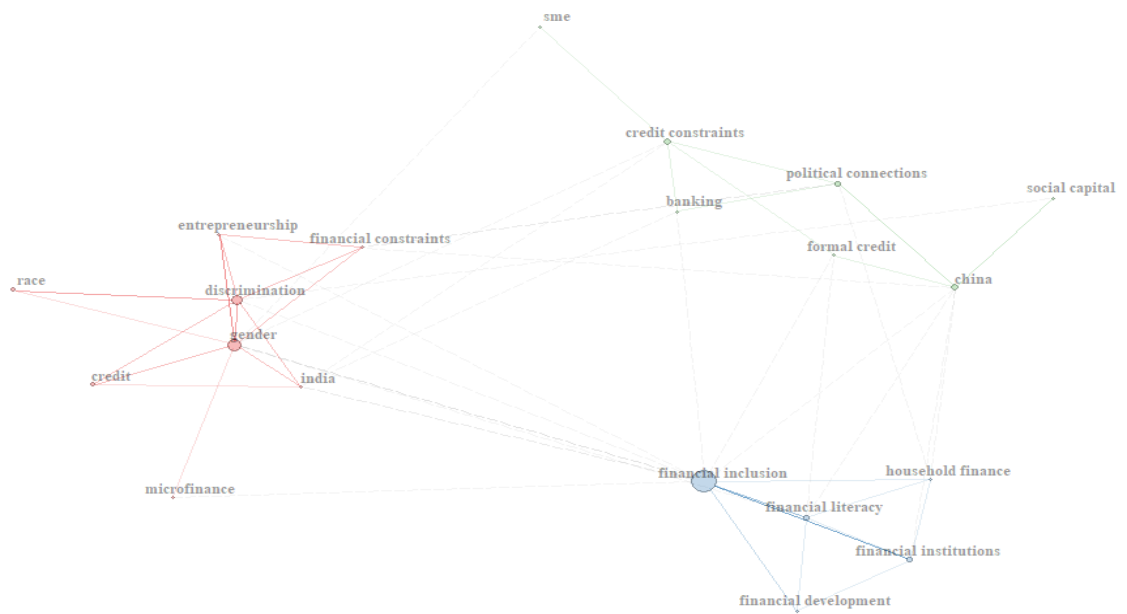
Figure 2.5: Identification of the most mentioned keywords in "determinants of finance" though bibliometric co-word analysis between 1985 and 2020.

After uploading bibliographic meta data-set in ⬢ Data, **(iii)**-***Data analysis:*** descriptive statistics can be seen under the ⬢ Data, 👤 Authors, and 👥 Citations. **(iv)**-***Data visualization:*** co-network analysis can be seen under the 🌲 Tree, 📖 Map, Ⓦ Words, ⬙ Thematic Map, and ⌃ Network tabs. Hence, in case Shinyapp disconnects from the server, "reload from the server" will appear, click on it, and repeat the process as mentioned above. Additionally, the guideline is presented in Appendix A.3.

2.8 Limitations of bibliometric study

"Any metric can be gamed, especially singular metrics such as citation counts."

- Christine L. Borgman

Bibliometric –as a method and discipline- has gained increasing attention especially by researchers, policymakers and research directors (Reuters, 2008) to evaluate research performance. Howbeit, it is also subjected to several limitations. As Laloë and Mosseri (2009) mentioned, the use of indices (i.e., h-index) is easy and an attractive way, but mostly unscientific. In the same way, publication count, a widely used indicator, is criticized, the quantity matters rather than quality[7], gratuitous co-authoring, and publication practices across fields (Schneider et al., 2016).

Garfield and Merton (1979) comprehensively answered the reasoning behind the citers' motivation for citing earlier works:

- Paying homage to pioneers

- Giving credit for related work (homage to peers)

- Identifying methodology, equipment, etc.

- Providing background reading

- Correcting one's own work

- Correcting the work of others

- Criticizing previous work

- Substantiating claims

- Alerting to forthcoming work

- Providing leads to poorly disseminated, poorly indexed, or uncited work

- Authenticating data and classes of fact-physical constants, etc.

- Identifying original publications in which an idea or concept was discussed

- Identifying original publications or other work describing an eponymic concept or term.

- Disclaiming work or ideas of others (negative claims)

- Disputing priority claims of others (negative homage).

Bavelas (1978) mentioned "the two extremes of this array of reasons might be true scholarly impact at the one end (e.g., significant use of the cited author's theory, paradigm, or method) and less-than-noble purposes at the other (e.g., citing the journal editor's work, or plugging a friend's publications)." Furthermore, it is possible that norms for citing vary from discipline to discipline.

There is a number of reasons why citing author did not cite other documents, as Kochen (1974) suggested "it is not surprising that there is a great deal of arbitrariness in the way authors select references for their bibliographies. Undoubtedly, many documents which should have been cited are missing and many cited documents are only slightly relevant."

Bibliometric analysis helps scholars to see the big picture, yet we encounter several additional limitations. We only consider articles in English. WoS data-base is comprehensive but not exhaustive. WoS data-base does not include all the journals, nor conference proceedings all around the world. This study does not exclude the self-citation[7] Another limitation, which is selecting the relevant articles for the meta-data, can be burdensome. Some articles may be excluded by the keyword selection or some recent quality articles may be overlooked due to lack of citations.

2.9   Related literature

This section reviews the possible future research streams that emerge from existing literature. The main papers, discussed in this section, are the most influential articles, affiliation, and journals (Table 2.6, Table 2.3, and Table 2.4), and the rest is the author's choice (content analysis).

---

[7] Hence, if an author several poor quality papers he/she is most likely to cited himself/herself which may cause the article has an influential effect on the literature. For this reason, we prefer to filter by SSCI and SCI-EXPANDED.

2.9.1 Access to finance and economic growth

Access to finance, usually defined as access to and use of financial services (Fungáčová et al., 2014; Hannig and Jansen, 2010), seeks to draw the involuntary excluded population into the formal financial system by expanding available financial supplies (Demirguc-Kunt and Klapper, 2013; Demirguc-Kunt et al., 2016). It is worth noting, access to and use of financial services are employed interchangeably. Yet, they refer to supply and demand (actual financial service consumption which is determined by demand and supply) of the financial services.

The usage of financial services, which benefits individuals, SMEs, and governments, has been associated with the stability of the economy and well-being. Particularly, access to external finance allows disadvantaged individuals to enhance opportunities such as education, health care, and investment (Demirguc-Kunt and Klapper, 2012), growth in existing, and new business opportunities while ensuring economic growth (Gan et al., 2014). Whereas, scarcity of financial products limits a range of services available to the governments, businesses, and households, which bars them from performing at their fullest potential and causes to diminished labor productivity, growth, income, and restrain the economic growth, as a consequence creates income inequity (Claessens, 2006; Demirguc-Kunt and Levine, 2008).

Regardless of the importance of external finance, a large number of studies have shown the determinants of accessing formal financial services, at both households and financial institution levels, depending on the applicants' characteristics. Especially in the developing countries those having favorable characteristics such as belonging to a major ethnic group, being male, having higher education, having high-income (Demirguc-Kunt and Klapper, 2013), being employed, holding assets (Campero and Kaiser, 2013; Claessens, 2006), residence in the urban area (Tejerina et al., 2007), and having political connections (Chen and Jin, 2017; Li et al., 2008) are more likely to access from formal financial sources.

Empirical findings pointed out that the use of financial services increases the amount of savings, consumption, and investment (Aportela, 1999; Ashraf et al., 2010; Dupas and Robinson, 2009). Yet only about 50% of adults around the world has access to bank accounts, banking penetration is even lower particularly in the developing countries (Demirguc-Kunt and Levine, 2008).

The importance of access to finance and financial development has drawn attention. Indeed, in

2010, accessing the formal financial services, for both households and enterprises, was regarded as one of the key pillars by G20 Summit as well as expanding and improving it (Bank, 2011).

A rising literature documents that finance contributes to "cause" of growth for countries, firms, and households (Beck et al., 2000; Demirguc-Kunt and Maksimovic, 1998) which increases income level, improves income distribution, and distributes opportunities more equitably which matters most to small borrowers (e.g., poor households, small firms), helps to reduce poverty (Beck et al., 2004; Xu et al., 2003), reduces child labor, increases the educational level and savings of the households.

Even in the time of financial distress in 2008, financially constrained companies had a series of financial problems compare to non-constrained firms. Campello et al. (2010) surveyed 1,050 Chief Financial Officers (CFOs) in the U.S., Europe, and Asia. He found that CFOs of the financially constrained firms are planned to have deeper cuts in employment, capital spending, and tech–spending. Chodorow-Reich (2014), using 2,000 non-financial U.S firms, Popov and Rocholl (2018), using more than 30.000 private and public firms in Germany, and Gerlach-Kristen et al. (2015), using 2565 Irish firms, showed that financially-constrained firms are more likely to layoff compared to non-constrained firms.

Financial constrains can be defined in 2-fold (i) quantity constraints[8] or (ii) higher loan rate Hayashi (1985). Additionally, financial constraints may cause a decrease in future investments (Gerlach-Kristen et al., 2015) and GDP growth (Bank, 2019), despite the advantages of finance, access to finance is far bellow the universal level and skewed toward the already better-off, particularly in developing countries.

2.9.2 Measuring financial inclusion

Access to finance is measured as access, usage, and the quality of the financial services presented as G20 Financial Inclusion Indicators. Financial inclusion can be measured by:

- The percentage of adults with a formal account,

- The percentage of adults with at least a formal loan,

- The percentage of SMEs with formal bank accounts,

---

[8]It is motivated by Stiglitz and Weiss (1981) and Jaffee and Russell (1976)

- The percentage of SMEs with an outstanding loan,

- The number of bank branches[9] and ATMs[10] per 100,000 adults.

These indicators will help policymakers to monitor the developments of financial inclusion for both nationally and globally.

Using Global Findex data with 123 countries and more than 124,000 individuals, Demirguc-Kunt et al. (2016) found that account ownership is around 91% in developed, 41% in developing, and 19% of adults in low-income economies. They found in developing countries 46% of men and 37% of women have formal account ownership, and those who have a formal account are equally use their account for saving. In this data-set, among 65,000 adults without any formal account, 66% of adults are reported they do not enough money, 24% of adults found the bank accounts expensive, 20% of adults mentioned banks are far away, 13% of adults do not trust banks, and 17% of adults do not have the necessary documentation. This study also showed that access to finance is associated with low banking costs and the physical distance of bank branches.

2.9.3 Determinants of finance

It is worth noting, access (refers to supply) to and use of financial services (refers supply and demand) are employed interchangeably, yet they refer to the different thing. Expanding access to finance should be integrated by understanding the geographic (e.g., lack of nearby bank branches) or socioeconomic (e.g., income, social, ethnic groups, .etc) barriers In comparison, use of financial services imply to actual service consumption Hannig and Jansen (2010). However, in literature access to finance, usually regarded as access to and use of financial services (Fungáčová et al., 2014; Hannig and Jansen, 2010), seeks to draw the involuntary excluded population into the formal financial system by expanding available financial supplies (Demirguc-Kunt and Klapper, 2013; Demirguc-Kunt et al., 2016).

Financial service providers not only find the poor households and small firms riskier but also costly to reach rural areas to provide a physical infrastructure where there might be a lack of security. They are usually favored to bigger companies & wealthy households and do not offer many solutions to small borrowers such as poor households and small firms. They may demand

---

[9]See at: http://data.imf.org/?sk=479a1de3-9ace-4c44-a97a-162d0bb3c9f5&hide_uv=1

[10]See at: http://data.imf.org/?sk=e5dcab7e-a5ca-4892-a6ea-598b5463a34c&sId=1460054136937

high minimum deposit, administrative fees, fixed cost loans, collateral, some favorable characteristic (e.g., male, white), and literacy requirement, which may cause to small borrowers (e.g., poor households and small firms) to be exclude from the formal financial services.

In loan market, discrimination occurs based on the participant characteristics' (such as gender, ethnicity, political affiliation, etc.) after controlling relevant factors of (e.g., credit history of the participant).

Discrimination creates barriers to accessing external funds and discourages entrepreneurs from implementing new ideas. Therefore, studying the existence of discrimination on personal attributes is essential to understand socioeconomic barriers.

2.9.3.1   Education

According to the growth theory (Romer, 1990; Solow, 1956), economical growth depends on the stock of technological progress, as well as physical and human capital (i.e., education). Financing at the individual or firm-level will have a direct effect on contributing to these things.

The study of Ellis and Lemma (2010) show that borrowing from formal financial services allows individuals to obtain a larger amount of money, which makes accessing formal financial services more desirable for investment purposes. Indeed, empirically showed the existence of a negative relationship between education and risk-aversion. The study of Riley Jr and Chow (1992) showed that risk tolerance increases with the level of education. Orens and Reheul (2013) showed that CEOs with higher level of education are receptive to investment opportunities and less concerned about the use of cash.

Individuals, with a high level of education, are more likely to access to formal financial services, save, and invest comparing the less educated individuals (e.g., households, self-employed). Using Vietnamese SMEs from 2004 to 2014, with 15,818 observations, Nguyen et al. (2020) found that the entrepreneur's with low-education levels are more likely to constraint and they are more likely to discouraged to access to formal financial services. Using Tanzanian SMEs data, Kira (2013) showed that access to finance ease with the firm's management education level. A number of literature has documented that access to formal financial services has a positive association with higher education.

Using 1988-1993 Consumer Expenditure Survey, the study of Grant (2007) shows that with a young college educated households are the most constrained.

2.9.3.2   Urban/Rural registration

Huge urban-rural income variation may cause great social and political problems which may affect economical and social stability negatively (Kanbur and Zhang, 1998). Based on the Demographic and Health Surveys (DHS) data, Young (2013) compared 65 countries and found that the income gap between rural & urban areas can be explained by 40% of the country's total income inequity, and this gap is even more for cross-country variation.

The study of Galiani and Schargrodsky (2010) shows, using data from Buenos Aires in 1989, land titling has a positive modest impact only accessing mortgage and has no impact on other forms of finance. Moreover, they found household investments increased when usufructuary property rights were given to the poor households. Bauer et al. (2016) discussed the importance of access to finance in the rural area, using 1338 households in Vietnam Access Resources Household Survey in 2012. The study showed that access to rural finance has created a remarkable poverty reduction in Vietnam and increased the total income of rural households.

2.9.3.3   Gender

The evidence shows that female and male-led firms are significantly different in terms of financial structure (Cesaroni, 2010). Usually, women-led firms are younger, smaller, show smaller marginal growth, and more likely to fail compering to male counterparts. Klapper and Parker (2011) analyzed a large body of literature on the relationship between gender and entrepreneurship. The study shows that women operate low capital industries which require less funding, and prefer less working hours since women have also family commitments. Furthermore, women-owned firms prefer to depend on their own capital rather than external financial sources, which causes the small profit in women-owned firms. As a result, financial providers are reluctant to lend money to women-led firms. Non the less, the existence of pure gender effect even after controlling the firm features is still an open-ended question.

According to the U.S Census Bureau, in the USA, there were 6.5 million women-owned firms which generated around 940$ billion and employed 7.1 million workers in 2002. Although the number of firms held by women increased by 20%, doubling the increase in US firms overall between the period from 1997 to 2002, in the same time span the revenues for the firms increased around 15%

and 22%, women-owned and US firms overall, respectively. The study of Robb and Wolken (2002); Rosa et al. (1996); Watson (2002) show that women-owned firms have low-level of employment, sales, and profits because they mostly concentrated in small competitive industries which makes easier to start new businesses, but difficult to sustain the business and they are likely to fail (Brush, 1992; Du Rietz and Henrekson, 2000; Fairlie and Robb, 2009; Loscocco et al., 1991).

Using four-waves of the Survey on the Access to Finance of Enterprises (SAFE) data on 24.000 SMEs which includes demand-supply effects (e.g., self-restraint or credit-ability), and firm characteristics (e.g., size, age, sector, owner of the firm, and the reason why the firm did not apply for loans) for the largest European countries (Germany, France, Italy, and Spain) which represents more than 99% of the European firms, Stefani and Vacca (2015) showed that women-led firms have a higher loan rejection rate and they do not apply for loans as frequently as their male counterparts, since they generally anticipate a rejection.

In addition, Barber and Odean (2001); Bellucci et al. (2010); Dohmen et al. (2005); Jianakoplos and Bernasek (1998) showed that women are more risk-averse. They demand less bank loans compare to their male counterparts and are more likely to sustain low debt-equity ratios.

Using the two waves (2004 and 2005) of Business Environment and Enterprise Performance Survey (BEEPS), implemented in 34 countries (Eastern-Western-Central Europe and Asia) and after removing the firms when the majority of the shareholders were represented by the government, legal person, general public, and keeping only individual own firms with a final sample of 5.022 SMEs, the empirical study of Muravyev et al. (2009) showed the evidence of gender discrimination after controlling relevant firm features (such as creditworthiness and performance of the firms). Female entrepreneurs have a 5.4% lower probability of accessing loans, and they pay 0.6% higher interest rate.

Using 7800 individually-owned small firms including 23 industry, Bellucci et al. (2010) found female-led firms are faced with more strict credit availability, and they need to provide 5.2% higher collateral comparing to their male counterparts.

After excluding incomplete records and government subsidies firms with a total of 1.209,078 observations, the study of Alesina et al. (2013) show women in Italy pay around 9 basis point on average more for credit compared to men. Yet, there is no evidence which indicates women are riskier than their male counterparts. This study also mentions women's roles in families. Family

ties are very strong in Italy among other OECD countries. There is a traditional husband-wife role, negatively affecting women labor force participation and the market activities (Alesina et al., 2013).

Cavalluzzo et al. (2002) used the 1993 National Survey of Small Business Finances (NSSBF) with a final sample of 4,570 SMEs, and found that women are reluctant to apply for credit because of the fear of denial as the lender concentration increases. Using Kauffman Firm Survey data-set, Coleman and Robb (2009) found that women start-up capitals and follow-on investments are significantly lower than men, and women prefer to rely on personal financial sources rather than external ones. Being less favorable to access to finance is not due to gender discrimination because of the characteristics of the firms.

Ghosh and Vinod (2017) show, using the Indian National Debt Survey with a total of 110,800 households, female-headed households are more constrain accessing finance compared with those headed by men.

Following the methodology by Beck and De La Torre (2006), the study of Asiedu et al. (2013) shows, using 34,342 non-agricultural firms with minimum five employees from 90 developing countries during 2011-2016, female-owned firms are more financially constrained than male-owned firms.

Using World Bank Global Findex data with total 141 countries, the study of Morsy (2020) shows that women are excluded from the financial services in countries where the share of foreign-owned banks are less and state-owned banks are more, credit information less opaque, and education gap between women and men is large.

2.9.3.4 Ethnicity

Using the SSBF data-set through 1993 and/or1998, Blanchard et al. (2008); Blanchflower et al. (2003); Cavalluzzo and Cavalluzzo (1998) found a significant dispersion on the loan approval rates based on the ethnicity of the applicants which can not be explained by the characteristics of the firms.

The benchmark model of Blanchflower et al. (2003), with only controls for race and gender, demonstrates that the denial rates for black-own firms 42.6% and 38% were higher than white-own firms in 1993 and 1998 respectively. And even after controlling for the creditworthiness of the firms, the denial rates for black-own firms were 27.7% and 30.1% higher than white-own firms in 1993 and

1998 respectively. Additionally, loan rejection rates were the highest for the Black firms compared to other ethnic/minor groups (e.g., Asian, Native American, Hispanic, and Female) for both years,

Using SSBF data-set, for the period 1998-1993, Blanchard et al. (2008) showed that the number of loan applicants for credit is almost the same for minority SME owners and white men. Although denial rates for Blacks are 42.4%, Hispanics 32%, and other races 20.7% more than white business owners. The interest rate paid for Blacks 0.65%, Hispanic 0.82%, and other races 1.06% are more comparing to white-male. Using 1990 Boston data, Munnell et al. (1996) estimated that Black and Hispanic mortgage loan applicants were 8.2% more likely to be rejected. Cavalluzzo et al. (2002) used the 1993 National Survey of Small Business Finances (NSSBF) with a final sample of 4,570 SMEs, and found that Africans are reluctant to apply for credit because of the fear of denial as the lender concentration increases. Even though, the study of Cavalluzzo and Wolken (2005), adding personal wealth as a predictor to extend the study of Cavalluzzo et al. (2002), founds that personal wealth can only explain around 10% of the different in denial rates between Hispanic- and Asian-owned firms compare to those owned by whites, whereas this proportion is much smaller to explain the different in denial rates between African-American and white firms.

Asiedu et al. (2012) studied how accessing finance is relevant to applicants' race, ethnicity, and gender following the methodology by Blanchard et al. (2008). They found that loan denial for the Black firms is about 22% and 40% higher than White male-owned firms for the years 1998 and 2003 respectively. Whereas denial rates for the other minority groups were relativity smaller.

2.9.3.5   The role of political connections

Studies have shown that there is a direct effect of accessing external finance which is shaped by the country's legal system. Allen et al. (2012) demonstrated that financial inclusion depends on the country's characteristics. Access to loan is higher in the countries where there is an efficient legal system, and political stability. Consequently, in countries with a weak justice system, firms struggle to access external financial services which diminish economic growth (Maksimovic and Demirgüç-Kunt, 1996; Porta et al., 1998). Securing property rights and enforcing contracts are inefficient (Frye and Zhuravskaya, 2000; Hay and Shleifer, 1998; McMillan and Woodruff, 1999). In those countries (e.g., Brazil, China, etc.), close ties with the current government or running political party helps businesses and firms to overcome market fluctuations.

Faccio (2006) found that firms having political connections create favorable access to finance which eventually strengthen the financial conditions and exaggerate the value of the firms (Fisman, 2001; Johnson and Mitton, 2003; Li et al., 2008; Ramalho, 2007). Claessens et al. (2008) studied the relationship between campaign financing and firms future political favors using the data from 1998-2002 Brazilian elections. They found that higher campaign contributions are associated with higher stock returns as well as higher bank leverage of firms.

Using 2400 enterprises across five regions and 18 cities in China, the study of Firth et al. (2009) shows that Chinese state-owned banks extend loans to better-governed and financially healthier firms.Granting loans and loan size is associated with the firm's profitability and political affiliation of the ownership.

Using 2011 Global Findex data-set, Fungáčová and Weill (2015) found formal account and formal savings in China is higher than other BRICS countries, yet accessing formal credit is only 7% in the last 12 months in China, the lowest comparing with BRICS countries. The reason why accessing formal credit so low is because formal credit is very concentrated, and directed to large state-owned firms (Geng and N'Diaye, 2012; Hale and Long, 2011). Indeed, granting loans in China covers around 1% of the country's SMEs. Cull et al. (2015) examined, using 2005 World Bank data covered 120 cities in Chinese with 12,400 firms, Chinese firms tend to face strict financial constraints if the CEO is non-government-appointed. Non-state Chinese firms heavily depend on their internal funds which severe the investment.

2.9.4 Policies to broaden financial inclusion

Access to finance remains a major issue, especially for developing countries. If loan applicants cannot fulfill the minimum financial requirements (e.g., income, financial assets, or size of the company), or do not have some favorable characteristics (e.g., being male, white, major ethnicity, employment, or having political connections), they are usually excluded or they are subject to pay higher fees.

Particularly, disadvantaged groups are excluded from access to finance (Shoji et al., 2012) due to the lack of income. Indeed, many small borrowers mentioned problems of demand collateral, high fixed cost, high administrative fees, and high rejection rate which creates income inequity (Claessens, 2006; Demirguc-Kunt and Levine, 2008).

An efficient way to reduce the income inequity, boost consumer spending, and ensure economic growth, is to swift government spending toward social welfare programs (Gan et al., 2014) considering social spending in China is 12.3% (social security not included) whereas 36.6% in the USA.

One way to finance the poorest of the poor could be microfinance. On the one hand, Honohan (2004) showed that microfinance penetration is useful for the poor but has no effects on poverty. On the other hand, Morduch and Haley (2002) showed that microfinance reduces poverty by mitigating credit constraints which reduces child labor and increases education. Drawing the "unbanked" population to financial services may seem difficult in the short run but in the long run, it will pay off as financial access reduces income inequity, diminishes the Gini coefficient (Honohan, 2008), and increases GDP growth (Beck et al., 2000; Levine, 2004).

Despite the importance of accessing formal loans, Beck et al. (2007); Demirguc-Kunt and Klapper (2012) demonstrated having access to a formal loan is far from the universal level in many developing countries, and the large world population do not have access to or use it (Demirguc-Kunt and Klapper, 2012; Research and Consulting, 2009). Yet particularly in developing countries, access to finance is limited, and skewed toward the already better-off (e.g., wealthier individuals or large enterprises).

Claessens (2006) suggested that accessing financial services is more important for financial institutions (especially smaller ones) comparing households. Particularly, new and small firms are often less established and have high exposures to market risks (e.g., financial crisis, macroeconomic volatility). Beck et al. (2005) showed that, with a sample of 99 countries including households, SMEs, and countries, more efficient banks and well developed financial system have wider banking sector penetration which leads to almost equally distributed deposit & loan use within banking clients. Additionally, easing access to finance increases individuals saving, enhances banks' deposits, financial stability, decreases the Gini coefficient, and inequity. On the other hand, the absence of access to finance affects growth, poverty reduction, and accumulative savings negatively.

Take into account, as Schumpeter (1942)[11] argued in the 7th chapter of his famous book "Capitalism, Socialism and Democracy" churn of innovation and competition brings a "creative destruction" which is the core of capitalism. The hypothesis of Gucz (2014) suggests that rising inequity

---

[11]Schumpeter (2013)

causes financial crisis, and reasoned the subprime crises (a mortgage was a good substitute of the income redistribution in contrast mortgage finance cause lending boom and housing prices run-up which increased the amount of the loan), and Kumhof et al. (2015) proved that the probability of a crisis increases when the low and middle income households have a higher consumption level leading accumulative loans. Identifying the main reasons of financial vulnerability to implement ex-ante policy implementations may help to sustain financial stability. As a consequence, officials developed Basel III and the European Banking Union to force banks both to manage their risk more prudently and to improve the banking supervision which is subject to regulation and capital requirements (BIS, 2011).

The bottom line is external finance improves income distribution through economic growth which reduces poverty, and it is especially important for the disadvantaged groups (e.g., individuals or households) and small businesses (Beck et al., 2008; Xu et al., 2003). Policies must target especially economically disadvantaged groups and smaller firms. Additionally, they should secure the financial system, which depends on better protection of legal system and property rights, to increase financial inclusion (Claessens, 2006).

To broaden access to financial services several regulatory, legal, and financial market reforms have come across to tailor and standardize financial inclusion. Policymakers imposed some rules to ease the finance for individuals, households, and SMEs.

Conclusion

This is the first reproducible bibliometric literature review on "determinants of finance" over the time span, using R. We examine 210 published English articles for bibliometric analysis, retrieved from Web of Science (WoS). In this study, we identify:

(1) influential aspects of the research stream (e.g., countries (Table 2.2), affiliations (Table 2.3), journals (Table 2.4), authors (Table 2.5), and articles (Table 2.6). The top three countries and affiliations publishing on determinants of finance literature are respectively USA, United Kingdom, China and World Bank, Georgetown Univ, Dartmouth Coll. Yet Netherlands and Harvard Univ are not on the top three in most influential countries and affiliations, there are more appreciated by the scholars.

We identify the top journals in three categories, namely, number of published articles (Article.No.), total local and global citations per year (TLC/t and TGC/t). Yet, in Table 2.4 "World Development" journal is in top three position, "Management Science", "Quarterly Journal Of Economics", and "Annual Review Of Sociology" are highly appreciated journals by the researchers.

We explore, on the one hand, most productive authors publishing on "determinants of finance": Beck T., Wyly Ek., and Agier I. On the other hand, the top three authors based on the TLC/t and TGC/t are respectively Allen F., Cavalluzzo Ks, Aterido R., and Beck T., Fernandes D., Khwaja Ai. Additionally, we show the trend papers in (Figure 2.2) and find the most trend paper is "Do Lenders Favor Politically Connected Firms? Rent Provision In An Emerging Financial Market" by Khwaja Ai, 2005.

(2) two research streams through co-citation (Figure 2.3) and co-word (Figure 2.5 and Table 2.8) analysis coupled with content analysis: (i) lending to small borrowers (ii) lending to big borrowers.

(3) co-author network (Figure 2.4) among the authors publishing on "determinants of finance".

(4) combining quantitative (bibliometric) and qualitative (content) analyses, we identify 13 research questions (Table 2.7). For the content analysis, we studied the most cited articles found by bibliometric analysis and some other articles that we found. Some of our research question can be retrieved from the top-cited articles (Table 2.6).

Additionally, we provide the code for this study in github[2] and create an interactive user-face using Shinyapp[3]. We present all the articles, in alphabetical order, that we used for the bibliometric analysis in Appendix A.1, distribution of authors' affiliations per each journal in Appendix A.2, and screenshot of the Biblio Shinyapp in Appendix A.3. We recommend repeating the citation analysis, and considering other databases (such as; Scorpus, Google Scholars, Science Direct, etc.) for future studies.

	Author(s) Name	Year	Title	TLC	TLC/t	TGC	TGC/t
1	Fernandes D; Lynch Jg; Netemeyer Rg	2014	Financial Literacy, Financial Education, And Downstream Financial Behaviors	2	0.29	439	62.71
2	Khwaja Ai; Mian A	2005	Do Lenders Favor Politically Connected Firms? Rent Provision In An Emerging Financial Market	14	0.88	866	54.12
3	Pager D; Shepherd H	2008	The Sociology Of Discrimination: Racial Discrimination In Employment, Housing, Credit, And Consumer Markets	4	0.31	689	53.00
4	Campbell Jy	2006	Household Finance	1	0.07	769	51.27
5	Claessens S; Feijen E; Laeven L	2008	Political Connections And Preferential Access To Finance: The Role Of Campaign Contributions	12	0.92	557	42.85
6	Beck T; Demirguc-Kunt A	2006	Small And Medium-Size Enterprises: Access To Finance As A Growth Constraint	8	0.53	626	41.73
7	Beck T; Demirguc-Kunt A; Maksimovic V	2008	Financing Patterns Around The World: Are Small Firms Different?	4	0.31	360	27.69
8	Hastings Js; Madrian Bc; Skimmyhorn Wl	2013	Financial Literacy, Financial Education, And Economic Outcomes	2	0.25	198	24.75
9	Houston Jf; Jiang Ll; Lin C; Ma Y	2014	Political Connections And The Cost Of Bank Loans	1	0.14	172	24.57
10	Allen F; Demirguc-Kunt A; Klapper L; Peria Msm	2016	The Foundations Of Financial Inclusion: Understanding Ownership And Use Of Formal Accounts	13	2.60	118	23.60

	Author(s) Name	Year	Title	TLC	TLC/t	TGC	TGC/t
1	Allen F; Demirguc-Kunt A; Klapper L; Peria Msm	2016	The Foundations Of Financial Inclusion: Understanding Ownership And Use Of Formal Accounts	13	2.60	118	23.60
2	Aterido R; Beck T; Iacovone L	2013	Access To Finance In Sub-Saharan Africa: Is There A Gender Gap?	16	2.00	83	10.38
3	Blanchflower Dg; Levine Pb; Zimmerman Dj	2003	Discrimination In The Small-Business Credit Market	31	1.72	237	13.17
4	Muravyev A; Talavera O; Schafer D	2009	Entrepreneurs' Gender And Financial Constraints: Evidence From International Data	20	1.67	145	12.08
5	Fungacova Z; Weill L	2015	Understanding Financial Inclusion In China	9	1.50	64	10.67
6	Cavalluzzo Ks; Cavalluzzo Lc; Wolken Jd	2002	Competition, Small Business Financing, And Discrimination: Evidence From A New Survey	27	1.42	186	9.79
7	Zins A; Weill L	2016	The Determinants Of Financial Inclusion In Africa	7	1.40	89	17.80
8	Alesina Af; Lotti F; Mistrulli Pe	2013	Do Women Pay More For Credit? Evidence From Italy	11	1.38	81	10.12
9	Ghosh S; Vinod D	2017	What Constrains Financial Inclusion For Women? Evidence From Indian Micro Data	5	1.25	31	7.75
10	Cavalluzzo Ks; Cavalluzzo Lc	1998	Market Structure And Discrimination: The Case Of Small Businesses	26	1.13	111	4.83

Note: The table is sorted by on sorted based on TLC/t and TGC/t, respectively.

Table 2.6: The most influential articles

Research Stream	Research Direction	Article
(i) Lending to small borrowers (in blue):		
1	What type of approaches to financial education are required?	Fernandes et al. (2014)
2	Is it possible to have more fare and equal financial system?	Dwyer (2018)
3	What kind of policies must target minor groups?	Carter et al. (2015)
4	What kind of policies should be implemented to ease the women access to finance?	Aterido et al. (2013); Ghosh and Vinod (2017)
5	Can house owning or land titling programs increase investment on education and helps to access to finance in the long run?	Galiani and Schargrodsky (2010)
6	How human capital is affecting approval of loan?	Bruns et al. (2008)
7	Comparing banking-relationship in develop and developing countries for lending decision making.	Ullah Bhuiyan and Shah (2011)
8	Why banking-relationship is more important for developing countries?	Ullah Bhuiyan and Shah (2011)
9	What is the financial accessibility today?	Author's suggestion
10	What is the role of digital finance?	Author's suggestion
11	How access to finance differs between countries where the higher education is free?	Author's suggestion
(ii) Lending to big borrowers (in red):		
12	How the stock return or debt reduction differs for the firms, having political ties?	Claessens et al. (2008)
13	How GDP and welfare of a country are affected negatively, when the companies are favor in access to finance?	Claessens et al. (2008)

Table 2.7: Research directions

	Research Stream	Keywords
(i)	Lending to small borrowers (in blue and red)	Financial inclusion, financial literary, financial development, financial institutions, household finance, race, credit, micro finance, India, gender, discrimination, entrepreneurship, and financial constraints.
(ii)	Lending to big borrowers (in green)	SMEs, credit constraints, banking, formal credit, political connections, China, and social capital.

Note: See Fig 2.5.

Table 2.8: Keywords under each stream

Chapter 3

Access to Credit: The Self-Employment Case in the Chinese Labor Market

3.1   Introduction

Financial inclusion, often defined as the use of formal financial services. The relevance of financial inclusion has a growing interest among policymakers and researchers worldwide and been extensively discussed in the literature. This is mainly because access to and use of wide range of formal financial services have been associated with well-being individuals as well as economic growth, and stability in the country (Claessens, 2006; Demirguc-Kunt and Levine, 2008). On the contrary, Limited access to and use of financial services may have a negative impact on financial stability for individuals (e.g., households, self-employed, and SMEs) and may create a hold back the economic growth as a consequence creates income inequity (Claessens, 2006; Demirguc-Kunt and Levine, 2008). Therefore, access to credit (external funds) allows disadvantaged groups (i.e., poor households) to enhance opportunities (e.g., education, health care, and investment) that otherwise would not be affordable.

Limited use of formal credit can be traced back to late 1970s, when the new economic reforms legislated in 1978, policies target the urban sector (Wan, 2008). Most of the major formal financial service providers (e.g., state-owned & commercial banks as well as rural credit cooperatives), providing external finance to vast number of Chinese households living in rural or poor areas (Hannig and Jansen, 2010), were closed (Sparreboom and Duflos, 2012). Meanwhile, the urban–rural gap was widening by excessive machinery production. Within this context, reducing income inequity and promoting access to finance has become a policy concern. Thereby, understanding the applicants' characteristics is crucial for ex-ante policy design which targets to promote financial inclusion.

A rising literature documents the firm and individual level bank loans with bank-level data,

but a few have focused on the household's financial inclusion using Machine learning techniques. The aim of this study is to understand the characteristics of Chinese households to access to credit and its type. This study differs from previous studies in the following ways. First, It focuses on predictive modeling rather than exploratory modeling, using machine learning methods and compares the predictive powers, to explain access to credit and its type, between machine learning models. Second, this study does not implement machine learning models only to whole data-set (CHFS)[1], but also splits the CHFS data-set in 4 different data splits. Namely, urban and rural (Urb & Rrl), education under high-school and above high-school (Educ.0 & Educ.1), not having party affiliation and having party affiliation (CPP.0 & CCP.1), female & male (Sex.0 & Sex.1). Third, it compares the predictive power between Benchmark and data-splits which may help to policy maker to understand characteristics of the Chinese households and give ability to compare the characteristics between the Benchmark and other splits. Moreover, we compare the different characteristics of the households for accessing into formal, informal, and both type of credit, using 3 different asset owning variables (net-worth, net-worth minus home equity (NW-HE), and liquid assets) since we have ample data support to capture. Our empirical results show that using benchmark (BchMk) data has the lowest predictive performance comparing to other data-splits and grouping the households based on the political affiliation (CPP.0 & CCP.1) has the best predictive power overall. Additionally, we show that using other data-splits and the BchMk can be misleading to understand the characteristics of the Chinese households.

We choose China for mainly four reasons. Firstly, China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018). Secondly, in 2005 self-employment accounted for 10% of the urban labor force, which means that at least 64 million people were self-employed in China with 40 million were rural-urban migrants and 24 million were self-employed urban residents (Cui et al., 2013). Only in 2015, China had generated 117 million self-employed jobs. Thirdly, China is the world's first-largest economy and has rapid economic growth over the past few decades. Yet, China is still considered a developing country, and millions are below the international poverty standards. Finally, the use of formal financial services (e.g., bank loans and credit cards) is far lower than other emerging economies (e.g., Brazil, India, Russia, and South Africa) (Fungáčová et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012).

---

[1]In this study, "CHFS", "Benchmark", and "BchMk" are used interchangeably.

In order to assess the limitations on accessing credit, it is essential to analyze the use of financial services. However, one of the most reliable and up-to-date data source, Global Findex, analyses bank-level data which mostly consider state-owned firms, and it has so limited observations since its main purpose is to compare financial inclusion across 148 countries (Fungáčová et al., 2014). For this reason, we choose 2015 China Household Financial Survey (CHFS) which focuses on Chinese household population with 133,183 individuals, including 37,289 household-head.

The remainder of the chapter is structured as follows. Section 3.2 explains the related literature review. Section 3.3 describes the CHFS data-set. Section 3.4 describes the methodology. Section 3.5 prints the descriptive statistics, and Section 3.6 presents the results.

3.2 Literature Review

Access to financial services and funding problems may have a negative impact on the financial sustainability for individuals (e.g., households, self-employed, etc.), growth, and income inequity (Claessens, 2006; Demirguc-Kunt and Levine, 2008). For a constant increase in the economic growth government's aim should be toward finance self-employed. External finance is a key ingredient for self-employed people to learn, grow, and expand the production capacity, irrespective of what they are producing whether agricultural goods or industrial goods or services. Yet for large multinational companies, it may not be the case as they can fund the growth, geographic expansion, and product development often through issuing bonds and other alternatives that are generally not available to self-employed individuals.

Baydas et al. (1994) examined the data of micro-entrepreneurs in Ecuador in 1990, in which they divided the borrowers among several groups. Firstly, those who were completely rejected which implies those who applied for a loan but not succeeded. Secondly, those who applied for a loan but obtained less amount of loan than demanded. Thirdly, those who got what they demanded and lastly non-applicants. They implemented multi-nominal logit models to estimate the probabilities for the respondents to be in one of the four groups. In several other studies, we find a similar approach. For instance, Zeller (1994) uses a similar approach and assesses the probabilities for the mentioned groups, but he further differentiates between formal and informal credit access in Madagascar in 1992. On the contrary, Yuan and Xu (2015) mentioned low income individuals

might not have equal access to formal credit and mostly they are excluded accessing formal credit sources (Shoji et al., 2012).

The effects of race, ethnicity, gender, and being a part of the ruling Chinese Communist Party (CCP) in determining the credit are important, but there are some other factors like credit schemes which are equally important in shaping credit access (Yuan and Xu, 2015). Mushinski (1999) used the data of rural households in Malawi and incorporated the credit limit variable to check the decision to participate in the credit programs along with access to the credit and he concluded the participation in certain credit programs significantly improves the access to the credit. However, the composition of collateral for instance type of household asset was more relevant for formal credit access than total value however, increasing the credit limit is moderately exploited by the borrowers.

Deng and Meng (2013); Dickson and Rublee (2000); Huang et al. (2013); Li et al. (2007, 2008) found having party affiliation with the Chinese communist party has a positive effect on economical outcomes which sequentially creates easy access to formal credit and increases the performance and the profitability of the firms (Faccio, 2006; Johnson and Mitton, 2003; Khwaja and Mian, 2005; Li et al., 2008; Ramalho, 2007).

Giné (2011) analyzed the mechanism underlying the access to the credit, concentrating on two important elements of rural credit. He argued the co-existence of government and other microfinance institutions with informal lenders. Similarly, Turvey and Kong (2010) found that approximately two-thirds of farm household debt holders borrow from their relatives or friends while reaming households had access to credit cooperatives.

A strand of literature investigates the determinant of personal attributes. There is a great amount of empirical evidence available on access to credit focused on the impact of ethnic, racial, and gender disparity. Cole and Wolken (1995) found evidence that male-owned businesses are more likely to access the credit in comparison to female counterparts. Blanchflower et al. (2003) documented the evidence of the significant role of race in loan approval rates. They found, after controlling the factors for credit worthiness, that black-owned companies are two times more likely to be denied the credit. Carter, SL and Anderson, Susan and Shaw, Eleanor (2001) concluded the evidence of gender bias in accessing the credit. They found gender characterization effects negatively to female entrepreneurs when accessing the credit. Using Heilongjiang Province Agriculture data

in 2008, Tang and Guo (2017) examined the farmers accessing credit in China's rural market and suggested promoting formal financial development and improving rural financial supply efficiency. Rui and Xi (2010) used China's local statistical bureaus'data on rural households data and showed that 71% of the rural households are faced credit constraints which has a negative impact on their welfare and the level of consumption.

3.3   Data Description

Household finance behavior is crucial to understand the macro-financial policy, corporate planning, and household financial decisions. To analyze the Chinese household credit used, we used the 2015 China Household Finance Survey (CHFS) which was nationally implemented by The Survey and Research Center for China Household Finance, established in 2010 as a non-profit research institution. CHFS documents the micro-level household finance from all over China which includes socioeconomic (e.g., savings, cash, etc.) and geographic (e.g., region, household registration, province, etc.) characteristics of the households.

The third wave of the CHFS household-level survey was implemented with a three-stage probability proportion to size (PPS) sampling performed. In the first stage of sampling, 29 provinces (except Hong Kong, Inner Mongolia, Macau, Tibet, Taiwan, and Xinjiang), 351 counties were selected. The second stage involves selecting committees/villages from the counties/cities chosen in the previous stage, and the last stage involves selecting households from the 1396 committees/villages selected in the earlier stage with a total sample size 37,289. In addition to that small and micro enterprises covered 28 provinces, 79 counties, and 234 towns with the sample size was 5497 households.

3.3.1   Dependent Variables

In the 2015 CHFS survey, participants were ask if they have any formal loan and/or if they engage in any other loan besides formal loan, as well as the purpose of the loan (Table 3.1) and the reasons for not having formal loan in Table 3.2.

Table 3.1 presents the column wise percentage of the total column for the purpose of the loan and access to multiple loans were allowed. For example, a household head may need a micro-enterprise

formal and education informal loan. So the households' heads obtain mainly formal loans from 45.78% credit cards, 39.73% need informal loan for housing and 27.7% medical purposes, and 47.39% need both loan for housing of the column total.

Where as, total (multiple counting were not allowed) shows the total number of formal, informal, both, and no loans, 3818 (11.65%), 4576 (13.96%), 1571 (4.79%), and 22800 (69.58%), respectively. The observations for total were counted as follows, $Formal = 1$ if the participant has only borrowed from formal institutions (e.g., banks) or had a credit card debt, $Informal = 1$ if the participant has only borrowed from informal organizations (e.g., money lenders) or network (e.g., family, friends), and $Both = 1$ if the participant has both formal and informal loans, and $No\,Loans = 1$ if the participant has non of the mentioned loans for any kind of purposes[2].

The dependent variables were assigned as follows, $access\,to\,credit = 1$ if respondent has at least one of the loans (e.g., formal, informal, or both) for any purposes, otherwise $access\,to\,credit = 0$.

Additionally, in the survey, the participants were asked the reason for not having loan. Table 3.2 shows that 37% of the households were reported that they do not need loan. 45% never applied loan and 37.3% of them believe the loan would not be approved. The main reasons of the loan rejection were with 28.85% not having guarantor, with 29.69% having low income, with 24.37% not being familiar with the loan officer, or with 23.81% not having collateral while only 4.76% of them has recorded having a bad credit history.

To remove the biasness of the calculation, the households who were reported as they do not need loan and have no informal loan were eliminated from the observation.

	Formal Loan	Informal Loan	Both Loan	No Loan
Microenterprise	1.66%	3.32%	35.69%	0.56%
Housing	36.6%	39.73%	47.39%	15.81%
Vehicle	7.96%	5.42%	5.51%	3.76%
Education	4.08%	11.27%	7.25%	20.44%
Medical	0.91%	27.7%	0%	19.98%
Credit card	46.78%	0%	0%	18.96%
Others	2%	12.57%	4.16%	20.5%
Column Total	6342	6167	1034	154722
Total	3818 (11.65%)	4576 (13.96%)	1571 (4.79%)	22800 (69.58%)

Note: Column total: column wise percentage was taken and access to multiple loan type is allow. Total: the total number of formal, informal, both, and no loan in our clean data-set. After cleaning the data we lost both loan for medical purposes.

Table 3.1: Purpose of Loans

---

[2]Note: If participant has no loan $No\,Loan = 1$ indicates that he/she has not access to loan $access\,to\,credit = 0$

	Reasons	Percent
a	*Does your household currently need loan? (n= 16635)*	
1	I do not need it	37%
2	I need it but I never applied	45%
3	I need it but I was rejected	13%
4	I need it and the application is pending	4%
b	*Why have not applied for loan? (Ask when a = 2) (n= 1214)*	
1	I do not know how to apply for a loan	19.11%
2	I do not think the application will be approved	37.3%
3	The application process is troublesome	21.84%
4	The interest rate is too high	15.72%
5	The repayment period of form is unrealistic	3.31%
6	I do not know the staff at the bank/credit union	8.6%
7	I do not have a guarantor or any collateral	13.56%
8	Worried that I am unable to pay	17.62%
9	Others	8.11%
c	*Why was the loan rejected? (Ask when a = 3) (n= 361)*	
1	Outstanding bank loan	2.24%
2	No guarantor	28.85%
3	Not familiar with the loan officer	24.37%
4	Low income, loan officer is worried I could not repay it	29.69%
5	No collateral	23.81%
6	Bad credit history	4.76%
7	Project is too risky	2.52%
8	Policy reasons	12.89%
9	Others	6.72%

Note: Data is available for only micro enterprise-related loans and multiple answers were allowed.

Table 3.2: Reasons for not having formal loan

3.3.2 Explanatory Variables

Variables were set up as follows gender (male = 1, female = 0), married (married = 1, otherwise = 0), employed (employed = 1, otherwise = 0), education (high school or higher = 1, otherwise = 0), party (affiliation with Chinese communist party = 1, otherwise = 0), household registration (HR) (urban = 1, rural = 0), region (west, east, and center), Fin.Knowledge (Fin.Knowledge = 1 when household's head took a financial class or had defined him/herself has having financial knowledge well, otherwise = 0), Fin.Inter (Fin.Inter = 1 when household's head is located 1-km arrange of a formal institution, otherwise = 0) and age, income, and 3 asset holding variable were coded as continues. We define the household asset holding assets, based on the previous studies (Brandolini et al., 2010; Caner and Wolff, 2004; Haveman and Wolff, 2004; Huang et al., 2013), and we measure households' asset holdings with three different continuous variables: net worth (the value of financial and non-financial assets minus liabilities), net worth minus home equity

	CCP.0	CCP.1
Formal	10% (2673)	18% (1145)
Informal	0.15 (4067)	8% (509)
Both	5% (1249)	5% (322)
No Loan	70% (18490)	69% (4310)

X squared $= 482.1952$ d.f. $= 3$ $p = 3.444258e - 104$

Table 3.3: Access to loan type based on party

(NW-HE), and liquid assets (cash and other easily cash-able assets). Definition of the variables and abbreviations can be found Appendix B.1 and Appendix B.2, respectively.

3.4 Methodology

We consider CHFS with 4 different data-splits based on the household's head registration (HR) type, education, political affiliation with the current Chinese communist party, and gender. The acronym we used for each data split "Urb & Rrl", "Educ.0 & Educ.1", "CPP.0 & CCP.1", and "Sex.0 & Sex.1", respectively. We build three-binary and three-multi-nominal logistic model, for $y = Access\,Loan$ and $y = Loan\,Type$, respectively, for the four data splits and for our "BchMk" (CHFS) data-set, using one of the three asset holds. To check the robustness of the models we set 3 machine learning models namely, bagging (BAG), random forest (RF), and boosting (GBM) as well.

We remove outliers if the observation has an influential effect on the slope and those who did not need loan and have no informal loan. Following James et al. (2013), we used 10 cv (cross-validation) based on 80/20 train/test sets. Using stratified random sampling, we divided the data-set into 80% where we set our models, then we compare the prediction results with the 20% of the data-set and to make sure we are not over-fitting we divided the data set into 10 folds and repeat the same process 10 times then we took the average of 10 folds to find the best data-set with the highest predictive power to examine the Chinese households.

3.4.0.1 Linear Classification Models

Linear classification models can not handle sparse or skewed predictors and there are sensible to outliers. Generally, in a non normally distributed data linear models perform poorly compare to machine learning models that known as black box. Yet, linear models can be useful for interpreting

the relationship between dependent and independent variables.

3.4.0.2 Generalized Logistic Model

Let $\boldsymbol{y}$ be a vector of response variable of accessing credit for each applicant n, such that $y_i = 1$ if the applicant-i has access to credit, and zero otherwise. Furthermore, let $\boldsymbol{x} = \{x_{i,j}\}$, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ be a matrix of n observations with p characteristics of the applicants. The log-odds can be define as

$$log(\frac{\pi_i}{1 - \pi_i}) \; = \; \beta_0 + \boldsymbol{x_i}\beta \; = \; \beta_0 + \sum_{i=1}^{p} \beta_i \boldsymbol{x}_i \tag{3.1}$$

where $\pi_i = P(y_i = 1|\boldsymbol{x_i})$, β_0 is the intercept, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of coefficients and $\boldsymbol{x_i}$ is the i-th row of $\boldsymbol{x}$.

3.4.0.3 Multi-nominal Logistic Model

Multi-nominal model is the generalized form of binary logistic model (3.1) and can be define as

$$\pi_i^h = P(y_i^h = 1|\boldsymbol{x_i^h}) \tag{3.2}$$

where h presents the class labels ("1-of-h") on the basis of an input vector x_j, in our case x_j is loan types ("Formal Loan", "Informal Loan", "Both Loan", and "No Loan"). Furthermore, $y_i^h = 1$ if the weight $\boldsymbol{w}$ of x_j corresponds to belong a class and $y_i^h = 0$ otherwise. For $i \in 1, ..., h$ and the weight vectors $\boldsymbol{w}^i$ corresponds to class i

We set $\boldsymbol{w}^h = 0$ and the parameters to be learned are the weight vectors $\boldsymbol{w}^i$ for $i \in \{1, ..., h-1\}$. And the class probabilities must satisfy

$$\sum_{i=1}^{h} P(y_i^h = 1|\boldsymbol{x_i^h}, \boldsymbol{w}) = 1$$

3.4.1 Supervised Learning: Ensemble Trees

Tree-based models, consist of one or more nested if-then statements for the predictors that partition the data, are preferred to be used for many real-life modeling problems. Because, they are easy to implement and interpretable, can effectively handle many types of predictors such as sparse, skewed, continuous, categorical, etc. without the need to pre-process them, do not require the predictors' relationship like linear regression model does, can effectively handle missing data, can implicitly conduct feature selection.

On the one hand, single tree models, If the relationship between predictors and the response cannot be adequately defined by rectangular regions that contain more homogeneous outcome values, will have larger prediction error than other kinds of models. On the other hand, ensemble methods that combine many trees into one model tend to have much better predictive performance than single trees.

3.4.1.1 Bagging

Decision trees suffer from high variance[3]. Bagging (shown in Algorithm 1) is an ensemble procedure which reduces the variance and increases the prediction accuracy of a statistical learning method by considering many training sets $(\hat{f}^1(x), \hat{f}^2(x), ..., \hat{f}^B(x))$ from the population. Since we can not have multiple training-sets, from a single training data-set, we can generate B different bootstrapped training data-sets $(\hat{f}^{*1}(x), \hat{f}^{*2}(x), ..., \hat{f}^{*B}(x))$ by each B trees and take a majority vote. Therefore, bagging for classification problem defined as

$$\hat{f}(x) = \operatorname*{argmax}_{k} \hat{f}^{*b}(x) \tag{3.3}$$

Algorithm 1 Bagging

1: **for** $i = 1$ to m predictions **do**
2:     From the original data-set, generate a bootstrap sample.
3:     Train an unpruned classification tree on this sample.
4: **end for**

---

[3]If we split the training data-set randomly into two parts and set a decision tree to both parts, the results might be quite different.

3.4.1.2 Random Forest

Rather than considering the random sample of m predictors from the total of p predictors in each split, random forest does not consider a majority of the p predictors, and considers in each split a fresh sample of m_{try} which we usually set to $m_{try} \approx \sqrt{p}$ (Breiman et al., 1984).

Random forests which de-correlate the trees by considering $m_{try} \approx \sqrt{p}$ show an improvement over bagged trees $m = p$. Random forest can be constructed as shown in Algorithm 2 (Breiman et al., 1984).

Algorithm 2 Random Forest

1: Select m, the number of models to build.
2: **for** $i = 1$ to m **do**
3:    Generate a bootstrap sample of the original data.
4:    Based on this sample, train a classification tree model.
5:    **for** each split **do**
6:       Randomly select k predictors out of P of the original predictors, where k $(< P)$.
7:       Select the best predictor among the k predictors and partition the data
8:    **end for**
9:    Use typical tree model stopping criteria to determine when a tree is complete.
10: **end for**

3.4.1.3 Gradient Boosting

Unlike bagging trees, boosting does not use bootstrap sampling, rather each tree is fit using information from previous trees. An event probability of stochastic gradient boosting model is given by

$$\hat{\pi}_i = \frac{1}{1 + exp[-f(x)]'} \tag{3.4}$$

where $f(x)$ is in the range of $[-\infty, \infty]$ and its initial estimate of the model is $f_i^{(0)} = log(\frac{\pi_i}{1-\pi_i})$, where $\hat{\pi}$ is the estimated sample proportion of a single class from the training set.

3.4.2 Parameter Estimation

3.4.2.1 Maximum Likelihood and AIC

Unknown regression coefficients are estimated based on the available training data using maximum likelihood. Let D is the given sample, and the coefficients (β) are chosen to maximize the likelihood

---

Algorithm 3 Gradient Boosting

1: Initialized all predictions to the sample log-odds: $f_i^{(0)} = log(\frac{\pi_i}{1-\pi_i})$
2: **for** *iteration* $j = 1...M$ **do**
3:     Compute the residual $z_i = y_i - \pi_i$.
4:     Randomly sample the training data-set.
5:     Train a tree model.
6:     Compute the terminal node estimates of the residuals.
7:     $r_i = \frac{\frac{1}{n}\sum_i^n (y_i - \pi_i)}{\frac{1}{n}\sum_i^n \pi_i(y_i - \pi_i)}.$
8:     Update the current model $f_i = f_i + \lambda f_i^{(0)}$, where$\lambda$ is a shrinkage parameter.
9: **end for**

---

function $L(\beta : D)$.

Likelihood function is given by

$$L(\beta : D) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \tag{3.5}$$

The estimator $\hat{\beta}$ is obtained by maximum likelihood estimator which satisfies

$$\hat{\beta} = \underset{\beta \in \Theta}{\operatorname{argmax}} \; \hat{L}_n(\beta; y)$$

where Θ is the parameter space, and y has a Bernoulli distribution ($y = 0$ or $y = 1$). After maximizing the $L(\beta : D)$ Akaike information criterion can be define as follows

$$\text{AIC} = 2k - 2\ln(\hat{L}) \tag{3.6}$$

where k is the number of estimated parameters in the model.

3.4.2.2   The Area Under the Curve (AUC)

Graphically, the Receiver operator characteristics (ROC) curve is a one by one square and plots the true positive rate ($TPR = number\,of\,false\,positives/total\,number\,of\,negatives$) against the false positive rate ($FPR = number\,of\,false\,positives/total\,number\,of\,negatives$) across all the possible thresholds.

AUC is the area under the ROC curve and summarizes the performance across all the possible

thresholds. Where $0 \leq AUC \leq 1$ and $AUC = 0.5$ is random guessing, $AUC = 1$ model is perfectly fit, and $AUC = 0$ model is always wrong.

3.5 Descriptive Statistics

Table 3.4 shows the summary statistics of the Chinese Households as a whole. After cleaning the CHFS survey, 76% of the respondents were male headed households and almost 86% of the households' head were married with an average age 54. The majority of the households (48%) are prefer to live in the east, and 19% of the households have a political affiliation with the current Chinese Communist Party. Only 32 % of the households had a high school or higher education. The average annual income was 69731.55 CNY (SD = CNY 175283.50) and our three asset owning variables were networth, NW-HE and liquid assets were CNY 745862.37 (SD = CNY 1603699.60), CNY 708689.99 (SD = CNY 1526658.33), 799474.99 (SD = CNY 1633306.20), respectively. Our primary analysis shows that having access to formal, informal, both (formal and informal) and having not access to any type of loan are 11.65%, 13.96, 4.79%, and 69.58%, respectively.

3.5.0.1 Political Dispersion

Table 3.12 compares predictive performance of the models and we can see the CCP.0 & CCP.1 data-split has a slightly higher predictive performance overall. Since CCP.0 & CCP.1 data split has the best predictive power to define access to and type of loan, we will examine the descriptive statistics for it.

Table 3.5 reports the characteristics of the households' heads based on their affiliation with the current Chinese Communist Party (CCP.1) or not (CCP.0), respectively, with the chi squared (χ^2) (for factor variables) and t-test (for continuous variables) of the two groups.

At a glance, one can see the mean of gross income and the three asset owning variables are significantly high for the households who has affiliation with the current Chinese Communist Party (CCP.1) comparing to those who has not (CCP.0). Another thing takes our attention is 56.94% has a education above high-school, 59.05% are employed, 11.61% had defined him/herself has financial knowledge, 78% lives in urban areas, and access to formal loan is 18.22% households who has affiliation with the current Chinese Communist Party (CCP.1). The table also shows the

Variable	Factor					
	0			**1**		
Gender	Female	23.72%		Male	76.27%	
Married	Others	13.99%		Married...........................	86.00%	
Employed	Others	35.89%		Employed	64.10%	
Education	Others	67.92%		High-school or above	32.07%	
Party	Others	80.81%		Party	19.18%	
HR	Rural	33.64%		Urban	66.35%	
Fin.Knowledge	Others	94.54%		Fin.Knowledge	5.45%	
Fin.Inter	Others	81.53%		Fin.Inter	18.46%	
Access Loan	Others............................	69.58%		Access Loan	30.41%	
Formal	Others	88.35%		Formal	11.65%	
Informal	Others	86.04%		Informal	13.96%	
Both	Others	95.21%		Both	4.79%	
No Loans	Others	30.41%		No loans	69.58%	

	West	East	Center
Region	27.25%	48.26%	24.48%

Continuous							
	Mean	SD	Median	Min	Max	Skew	Kurt
Age	53.90	14.28	53.0	17.0	101	0.02	-0.50
Income	69731.55	175283.50	41750.0	-800000.0	5000000	16.70	396.64
Networth	745862.37	1603699.60	283353.3	-627904.2	19999748	6.50	58.35
NW-HE	708689.99	1526658.33	270636.5	-3614776.4	19999748	6.53	59.59
Liquid Assets	799474.99	1633306.20	325099.6	0.0	20000000	6.45	57.53

Note: HR stands for Household Registration. NW-HE is net-worth minus home equity. All the asset variables (e.g. income, net-worth, NW-HE, and liquid assets are in Chinese renminbi (CNY).

Table 3.4: Summary statistics

Variable	Factor		
	CCP.1	CCP.0	(χ^2)-test
Gender	80.31%	75.32%	p-value < 2.2e-16
Married	89.44%	85.19%	p-value < 2.2e-16
Employed	59.05%	65.31	p-value < 2.2e-16
Education	56.94%	26.17%	p-value < 2.2e-16
HR	78.05%	63.58%	p-value < 2.2e-16
Fin.Knowledge	11.61%	3.99%	p-value < 2.2e-16
Fin.Inter	17.05%	18.8%	p-value $= 0.001372$
Region- East	52.31%	47.3%	p-value $= 9.402$e-13
Region-Center	25.82%	27.6%	p-value $= 0.004385$
Region-West	21.87%	25.1%	p-value $= 8.985$e-08
Formal	18.22%	10.09%	p-value < 2.2e-16
Informal	8.1%	15.36%	p-value < 2.2e-16
Both	5.12%	4.72%	p-value $= 0.1761$
No Loans	68.57%	69.83%	p-value $= 0.05022$
	Continuous		
	CCP.1	CCP.0	t-test
Age	55.72(15.01)	53.46(14.07)	2.26 (0.21)***
Income	93097.17 (189449.94)	64184.65 (171285.19)	28912.52 (2611.07)***
Networth	1102972.99 (1941971.67)	661085.85 (1499929.43)	441887.15 (26170.78)***
NW-HE	1040123.66 (1851839.12)	630009.07 (1427490.00)	410114.588 (24950.02)***
Liquid Assets	1169314.39 (1980845.86)	711676.72 (1526257.28)	457637.67 (26686.67)***

Note: HR stands for Household Registration. NW-HE is net-worth minus home equity. All the asset variables (e.g. income, net-worth, NW-HE, and liquid assets are in Chinese renminbi (CNY).

Table 3.5: Summary statistics

percentage of the existence of financial intermediaries (Fin.Inter[4]) are in 1km range for the both groups, access to formal loan for is almost double for CCP.1 (18.22%) where as it is 10% for CCP.0

3.5.1 Analyzing Association Between Variables

Analyzing the association between pair of variables using the Chi-square (χ^2) test of independence. The Chi-Square (χ^2) test of independence is a non-parametric test and determines whether there is an association between categorical variables (i.e., whether the variables are independent or related).

3.5.1.1 Access to Loan and Loan Type

Table 3.6 shows the households column percentage of access to loan and its type based on Fin.Knowledge and Fin.Inter. On the one hand, we can see that household's head who has financial knowledge (Fin.Knowledge1) has 35% of access to formal loan where as this percentage can be dropped till 10% if the household's head has no financial knowledge (Fin.Knowledge0). On the other hand, we see

---

[4]Note: In the survey, the number of the formal financial institutions in 1km range is not provided.

that even the household's head has 1 km range from the formal financial intermediates (Fin.Inter1) or not (Fin.Inter0) the household's head access to formal loan is 13% or 11%, respectively.

The Chi-squared test (χ^2) shows a significant relationship between Fin.Knowledge and access to loan type with $p <$0.05 that indicates financial knowledge play a role to access to loan type. Where as there is no association between Fin.Inter and access to loan type is not statistically significant ($p = 0.07636063 \not< 0.05$).

	Fin.Knowledge0	Fin.Knowledge1	Fin.Inter0	Fin.Inter1
Formal	10%	35%	11%	13%
Informal	14%	5%	14%	13%
L.Both	5%	9%	5%	5%
No.Loan	71%	50%	70%	69%
Column Total	30979	1786	26716	6049

Note: for Fin.Knowledge: X-squared = 1180.502 d.f. = 3 $p < 2.2e - 16$, Fin.Inter X-squared = 6.864018 d.f. = 3 $p = 0.07636063$.

Table 3.6: Access to loan type based on Fin.Knowledge and Fin.Inter

3.5.1.2 Pearson Correlation

Table 3.7 shows the Pearson correlation between income and our 3 asset owning variables for all the data-splits and the benchmark model. All the correlations are under 50% which suggests we do not have to concern about the correlations between 3 assets variable and income.

	BchMk	Urb	Rrl	Educ.0	Educ.1	CCP.0	CCP.1	Sex.0	Sex.1
Income & Networth	0.45***	0.39***	0.48***	0.35***	0.46***	0.42***	0.46***	0.45***	0.43***
Income & NW-HE	0.43***	0.33***	0.46***	0.33***	0.41***	0.36***	0.43***	0.39***	0.38***
Income & Liquid Assets	0.46***	0.40***	0.51***	0.38***	0.47***	0.44***	0.49***	0.46***	0.45***

Note: Where the significance of the relationship is $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3.7: Pearson correlation

3.6 Results

3.6.1 Predictive power comparison

3.6.1.1 Robustness check

Table 3.12 prints out AUC of all the models. Where columns for $y = Access\,Loan$ and $y = Loan\,Type$ prints out the predictive power of each model for the relevant outcome y. One can

conclude that all the models shows similar predictive power yet GBM for the CCP.0 & CCP.1 data-split shows slightly higher predictive power comparing to other models for the both y.

Table 3.12 can be interpreted as follows, to calculate $y = Access\,Loan$, GLM, BAG, RF, and GBM models and to calculate $y = Loan\,Type$, MLM, BAG, RF, and GBM were set using the 3 different financial asset holding variables, with Network, NW-HE, and Liquid Asset as predictor, 1, 2, and 3. Overall, AUC of GBM models are higher across each model and data-set. AUC of the CCP.0 & CCP.1 split is 0.725 in GBM.2 and 0.733 in GBM.2 and GBM.3 models, for $y = Access\,Loan$ and $y = Loan\,Type$, respectively.

3.6.2 Generalized Logistic Models

3.6.2.1 Determinants of Credit Use

Table 3.12 presents the predictive power of each model used. We can see the model prediction for CCP.0 & CCP.1 data-split is similar to each other and relatively high comparing to other models with AUC is 0.708, 0.706, and 0.708 for GLM.1, GLM.2, and GLM.3. To avoid any confusion, we will compare only GLM.3 with BchMk.3 for $y = Access\,Loan$.

Table 3.8 presents the results of logistic models that regressed three different sets of asset variables on credit use for each data-splits and our benchmark. All the models regressed 11 variables with one asset holding variable each time: net-worth (GLM.1), NW-HE (GLM.2), and liquid asset (GLM.3).

As mentioned above CCP.0 & CCP.1 data-split has a higher predictive power to explain the characteristics of Chinese households. The interpretation of the Table 3.8 is GLM.3 for the BchMk data-set, being married (OR $= 1.25 \simeq 55\%$, $p < 0.05$), being employed (OR $= 1.27 \simeq 56\%$, $p < 0.05$), having a high school or above education (OR $= 1.14 \simeq 53\%$, $p < 0.05$), Fin.Knowledge (OR $= 1.69 \simeq 63\%$, $p < 0.05$), income (OR $= 1.11 \simeq 53\%$, $p < 0.05$), liquid assets (OR $= 1.21 \simeq 55\%$, $p < 0.05$), these explanatory variables shows a positive and significant association to access to credit.

The association and the significance between y and x does not change overall between BchMk and the different data-splits. Yet, GLM.3 for the CCP.0 & CCP.1 data-set, Fin.Knowledge OR $= 1.76 \simeq 64\%$, $p < 0.05$ and OR $= 1.39 \simeq 58\%$, $p < 0.05$ for CCP.0 and CCP.1, respectively. Increasing

financial knowledge has a positive affect of access to credit, where as financial intermediaries are positively associated with y yet it is not significant.

In short, to explain accessing credit ownership ($y = Access\,Loan$) can be interpreted differently base on the selected data-set, and selecting the wrong data-split may lead to misleading interpreting (in terms of the magnitude of the independent variables). Using not explanatory but predictive modeling may help policymakers to evaluate the characteristics of the Chinese household more accurately.

3.6.3   Multinominal Logistic Models

3.6.3.1   Determinants of Credit Type Use

Table 3.12 shows the predictive power of all the models for 3 different asset owning for each data-splits and our benchmark. All the models regressed 11 variables with one asset holding variable each time: net-worth, NW-HE, and liquid asset, as predictor, 1, 2, and 3, respectively. The multinominal logistic models for the CCP.0 & CCP.1 data-split have slightly higher predictive power comparing the other models with AUC 0.720, 0.717, and 0.720 for MLM1, ML2, and MLM3, respectively. Where the value of the dependent variable ($y = Loan\,Type$) is having formal (Fml), informal (Infm), or both (Both) loans are set as factor and compared with the based which is no loan $= 0$.

Table 3.9, 3.10, and 3.11 print out the results of the multinominal logistic models that regressed three different sets of asset variables on credit type used for each data-splits and the benchmark. All the models regressed 11 variables with one asset holding variable each time: net-worth (MLM.1), NW-HE (MLM.2), and liquid asset (MLM.3). To avoid any confusion, we will compare only MLM.3 with BchMk.3 for $y = Access\,Type$.

The interpretation of the Table 3.11 is to access to formal loan in MLM.3 for the BchMk.3 data-set, being married (RRR $= 1.93 \simeq 66\%$, $p < 0.05$) being employed (RRR $= 1.61 \simeq 62\%$, $p < 0.05$), having a high school or above education (RRR $= 2.07 \simeq 67\%$, $p < 0.05$), Fin.Inter (RRR $= 1.22 \simeq 55\%$, $p < 0.05$), Fin.Knowledge (RRR $= 2.07 \simeq 68\%$, $p < 0.05$), income (RRR $= 1 \simeq 50\%$, $p < 0.05$), liquid assets(RRR $= 1 \simeq 50\%$, $p < 0.05$), these explanatory variables shows a positive and significant association to access to type of credit.

As mentioned above CCP.0 & CCP.1 data-split has a higher predictive power to explain the

characteristics of Chinese households. The association and the significance between y and x does not change overall. Yet, MLM.3 for the CCP.0 & CCP.1 data-set, Fin.Inter ($RRR = 1.10 \simeq 52\%$, $p < 0.05$ and $RRR = 1.23 \simeq 55\%$, $p < 0.05$), Fin.Knowledge($RRR = 1.57 \simeq 61\%$, $p < 0.05$ and $RRR = 2.14 \simeq 68\%$, $p < 0.05$) for CCP.0 and CCP.1, respectively. Increasing financial knowledge has a positive affect of access to type of credit.

All the variables except gender, age, and region were found to have a positive statistically significant relationship to access to formal loan. In short, to explain accessing credit type ($y = Access\,LoanType$) can be interpreted differently base on the selected data-set, and selecting the wrong data-split may lead to misleading interpreting (in terms of the magnitude of the independent variables). Using not explanatory but predictive modeling may help policymakers to evaluate the characteristics of the Chinese household more accurately.

Conclusion

This study is to examine the characteristics of the Chinese households to access to credit and its type. We split the data-set into 4 different way (Urb & Rrl, Educ.0 & Educ.1, CPP.0 & CCP.1, and Sex.0 & Sex.1) and we compare the characteristics of the access to credit and its type between these data-sets, using 3 different asset owning variables (net-worth, NW-HE, and liquid assets), we built total 120 models, three logistic and multinominal logistic for four data-splits and BchMk, total 30 linear models, and to three machine learning (e.g., boosting, random forest, and bagging) models for four data-splits and BchMk, total 90 ml models, to explain characteristics of Chinese household for both access to loan and its type. The results are robust. And using not explanatory but predictive modeling may help policymakers to evaluate the characteristics of the Chinese household more correctly.

Our empirical results show that splitting the CHFS survey according to the political affiliation (CCP.0 & CCP.1) has higher predictive power and performs better to explains the Chinese household characteristics and most importantly using other data-splits and benchmark data-set can be misleading in terms of magnitude which affects the probability of accessing loan and the loan type.

The findings of this study line with Chen and Jin (2017); Fungáčová et al. (2014), and indicate credit ownership is low. Formal financial inclusion is particularly constrained which is distributed

economically advantaged groups as a consequence disadvantage households can not contribute to economic development. Also, households, having an affiliation with the current Chinese communist party, benefit from the use of formal loans. The formal loan is given mostly to state-owned companies rather than disadvantaged groups (e.g., individuals, households) (Fungáčová et al., 2014; Shoji et al., 2012) and political affiliation can help to access loan (Faccio, 2006; Khwaja and Mian, 2005).

The findings suggest that financial inclusion remains a major issue since loan applicant usually excluded if they can not full fill the minimum requirements (e.g., income, employment, and financial assets) which cause inequity to arise. An efficient way, to reduce income inequity, boost consumer spending, and ensure economic growth, is to swift government spending toward social welfare programs (Gan et al., 2014). Additionally, increasing the number financial intermediates and the household financial knowledge can help to access to formal loan. Policies must target economically disadvantage households to improve the financial inclusion.

Nonetheless, there are several limitations with CHFS data-set. Firstly, credit used was reported based on the household level, and unbalance formal loan accessibility restricts our study. Secondly, the majority of the households were male-headed which does not reveal the real gender biases to access credit. Lastly, ethnicity, and the number of financial intermediates which are closed to household were not reported.

Variables	BchMk	Urb	Rrl	Educ.0	Educ.1	CCP.0	CCP.1	Sex.0	Sex.1
Panel A:	**GLM(1)** - Model with Networth								
Gender	0.95	0.92	0.97	1.00	0.86***	0.97	0.97	——	——
Marital Status	1.25***	1.32***	1.33***	1.16**	1.79***	1.32***	1.36**	1.14*	1.41***
Age	0.58***	0.70***	0.53***	0.64***	0.48***	0.62***	0.43***	0.52***	0.58***
Employed	1.27***	0.92	1.40***	1.11**	1.75***	1.17***	1.40***	1.23***	1.25***
Education	1.14***	0.89	1.14***	——	——	1.10**	1.22**	1.21***	1.11***
Party	1.05	0.96	1.09**	0.87**	1.26***	——	——	1.12	1.05
HR	0.75***	——	——	0.72***	0.92	0.75***	0.78***	0.62***	0.79***
Region-East	0.64***	0.61***	0.66***	0.60***	0.74***	0.65***	0.72***	0.66***	0.62***
Region-Center	0.84***	0.87**	0.84***	0.87***	0.83**	0.86***	0.92	0.90	0.83***
Fin.Inter	1.02	1.06	1.03	1.07	0.94	1.04	1.01	1.06	1.02
Fin.Knowledge	1.70***	1.52**	1.66***	1.50***	1.55***	1.77***	1.39***	1.52***	1.72***
Income	1.12***	1.04	1.14***	1.05***	1.28***	1.10***	1.16**	1.15***	1.12***
Networth	1.20***	1.11***	1.26***	1.10***	1.33***	1.15***	1.44***	1.26***	1.19***
Observations	26,212	8,819	17,394	17,806	8,408	21,184	5,029	6,220	19,994
Log Likelihood	−14,794.52	−5,294.76	−9,433.94	−9,864.52	−4,761.02	−12,158.90	−2,600.08	−3,328.19	−11,390.92
Panel B:	**GLM(2)** - Model with NW-HE (Net-Worth minus Home Equity)								
Gender	0.94*	0.91	0.96	1.00	0.86***	0.97	0.96	——	——
Marital Status	1.27***	1.33***	1.35***	1.17***	1.83***	1.33***	1.39**	1.15**	1.42***
Age	0.58***	0.70***	0.53***	0.64***	0.48***	0.62***	0.43***	0.52***	0.58***
Employed	1.26***	0.92	1.40***	1.11**	1.75***	1.17***	1.39***	1.22***	1.25***
Education	1.16***	0.89	1.17***	——	——	1.11***	1.25***	1.23***	1.13***
Party	1.06	0.97	1.10**	0.87**	1.27***	——	——	1.13	1.06
HR	0.76***	——	——	0.73***	0.95	0.76***	0.80**	0.63***	0.80***
Region-East	0.66***	0.62***	0.68***	0.61***	0.77***	0.66***	0.75***	0.68***	0.63***
Region-Center	0.84***	0.87**	0.84***	0.86***	0.83**	0.86***	0.91	0.90	0.83***
Fin.Inter	1.03	1.06	1.04	1.08*	0.95	1.05	1.02	1.07	1.03
Fin.Knowledge	1.74***	1.55***	1.71***	1.53***	1.60***	1.81***	1.42***	1.55***	1.77***
Income	1.16***	1.06*	1.20***	1.07***	1.39***	1.14***	1.24***	1.21***	1.17***
NW-HE	1.10***	1.06**	1.14***	1.05***	1.16***	1.07***	1.25***	1.15***	1.09***
Observations	26,212	8,819	17,394	17,806	8,408	21,184	5,029	6,220	19,994
Log Likelihood	−14,834.94	−5,299.74	−9,474.36	−9,874.37	−4,788.65	−12,183.12	−2,619.91	−3,341.38	−11,418.90
Panel C:	**GLM(3)** - Model with Liquid Assets								
Gender	0.95	0.92	0.97	1.00	0.86**	0.97	0.97	——	——
Marital Status	1.25***	1.31***	1.33***	1.16**	1.79***	1.32***	1.36**	1.14*	1.41***
Age	0.58***	0.70***	0.53***	0.64***	0.48***	0.62***	0.43***	0.52***	0.58***
Employed	1.27***	0.92	1.40***	1.11**	1.75***	1.17***	1.40***	1.23***	1.25***
Education	1.14***	0.88	1.14***	——	——	1.09**	1.21**	1.21***	1.10**
Party	1.05	0.96	1.09**	0.87**	1.25***	——	——	1.11	1.05
HR	0.74***	——	——	0.72***	0.91	0.75***	0.78***	0.62***	0.78***
Region-East	0.64***	0.61***	0.66***	0.60***	0.74***	0.65***	0.72***	0.66***	0.62***
Region-Center	0.84***	0.87**	0.85***	0.87***	0.83**	0.86***	0.92	0.90	0.83***
Fin.inter	1.02	1.06	1.03	1.07	0.94	1.04	1.01	1.06	1.02
Fin.Knowledge	1.69***	1.49**	1.65***	1.49***	1.55***	1.76***	1.39***	1.52***	1.71***
Income	1.11***	1.03	1.14***	1.05**	1.27***	1.10***	1.15**	1.14***	1.11***
Liquid Assets	1.21***	1.14***	1.28***	1.11***	1.36***	1.17***	1.47***	1.28***	1.20***
Observations	26,212	8,819	17,394	17,806	8,408	21,184	5,029	6,220	19,994
Log Likelihood	−14,785.80	−5,290.73	−9,427.63	−9,861.19	−4,756.70	−12,152.80	−2,597.08	−3,325.96	−11,384.15

Note: Asset holding variables Networth, NW-HE (net-worth minus home equity), and liquid assets. Odd ratios (OR) are reported with significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. The number of the observation is approximately 80% of the total observation and the total number of observation may be vary based on the data-split group. The variables and abbreviations can be found in Appendix B.1 and Appendix B.2, respectively.

Table 3.8: GLM models

Variables	Fml	Infm	Both	Fml	Infm	Both	Fml	Infm	Both
		BchMk			Urb			Rrl	
Gender	0.84***	1.11***	0.97***	0.81***	1.20***	1.00***	1.16***	0.92***	0.84***
Marital Status	1.94***	1.15***	1.93***	1.82***	1.11***	2.27***	2.31***	1.15***	1.73***
Age	0.95***	0.98***	0.95***	0.94***	0.98***	0.94***	0.97***	0.97***	0.96***
Employed	1.61***	1.07***	1.57***	1.61***	1.15***	1.55***	1.11***	0.87***	1.13***
Education	2.08***	0.62***	1.16***	2.16***	0.57***	1.26***	1.33***	0.84***	0.90***
Party	1.44***	0.78***	1.13***	1.44***	0.73***	1.14***	1.26***	0.81***	1.19***
HR	1.52***	0.65***	0.72***	——	——	——	——	——	——
Region-East	0.77***	0.66***	0.56***	0.82***	0.64***	0.55***	0.64***	0.61***	0.57***
Region-Center	0.64***	1.03***	0.76***	0.70***	1.07***	0.75***	0.49***	0.98***	0.79***
Fin.Inter	1.23***	0.98***	1.10***	1.22***	0.91***	0.98***	1.23***	0.99***	1.12***
Fin.Knowledge	2.08***	0.80***	2.11***	2.00***	0.78***	1.78***	1.99***	1.03***	3.12***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
Networth	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00**	1.00***
Observations		26212			17393			8818	
AIC		42,266.04			27,160.73			14,952.09	
		Educ.0			Educ.1			CCP.0	
Gender	0.95***	1.12***	0.97***	0.81***	1.05***	0.84***	0.87***	1.17***	0.98***
Marital Status	1.80***	1.04***	1.49***	2.06***	1.13***	2.83***	1.45***	1.16***	2.53***
Age	0.95***	0.97***	0.95***	0.94***	0.99***	0.95***	0.94***	0.97***	0.94***
Employed	1.14***	0.99***	1.42***	1.95***	1.29***	1.94***	1.72***	1.20***	2.30***
Education	——	——	——	——	——	——	1.97***	0.73***	1.28***
Party	1.33***	0.77***	1.06***	1.58***	0.77***	1.24***	——	——	——
HR	1.32***	0.67***	0.68***	2.57***	0.43***	0.87***	1.77***	0.51***	0.58***
Region-East	0.74***	0.61***	0.56***	0.82***	0.81***	0.53***	0.95***	0.54***	0.58***
Region-Center	0.61***	1.04***	0.83***	0.71***	1.16***	0.69***	0.90***	0.97***	0.84***
Fin.Inter	1.23***	1.02***	1.11***	1.18***	0.72***	0.90***	1.10***	0.77***	1.17***
Fin.Knowledge	2.41***	0.82***	3.19***	1.84***	0.87***	1.86***	1.58***	0.57***	2.00***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00*	1.00***
Networth	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00**	1.00***
Observations		17804			8407			21183	
AIC		27,829.18			14,294.98			7,688.11	
		CCP.1			Sex.0			Sex.1	
Gender	0.84***	1.15***	0.97***	——	——	——	——	——	——
Marital Status	1.94***	1.13***	1.95***	1.59***	1.00***	1.63***	2.07***	1.23***	2.85***
Age	0.95***	0.98***	0.95***	0.94***	0.97***	0.96***	0.95***	0.98***	0.94***
Employed	1.57***	1.01***	1.38***	1.37***	1.03***	1.38***	1.80***	1.09***	1.46***
Education	2.04***	0.65***	1.07***	2.17***	0.57***	1.38***	2.09***	0.62***	1.14***
Party	——	——	——	1.43***	0.86***	1.06***	1.48***	0.79***	1.14***
HR	1.50***	0.66***	0.80***	1.99***	0.52***	0.65***	1.58***	0.69***	0.74***
Region-East	0.78***	0.66***	0.51***	0.71***	0.71***	0.54***	0.76***	0.60***	0.57***
Region-Center	0.60***	1.01***	0.72***	0.60***	1.12***	0.80***	0.64***	1.01***	0.79***
Fin.Inter	1.23***	1.02***	1.07***	1.32***	0.97***	1.01***	1.18***	0.96***	1.08***
Fin.Knowledge	2.18***	0.94***	2.38***	1.71***	0.66***	1.74***	1.88***	0.94***	2.08***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1	1.00***	1.00***	1.00***
Networth	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
Observations		5028			6219			19992	
AIC		34,686.78			9,462.19			32,711.43	

Note: Multinominal Logistic Model with Networth as Predictor. Relative Risk Ratios (RRR) are reported with significance: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. The variables and abbreviations can be found in Appendix B.1 and Appendix B.2, respectively.

Table 3.9: MLM.1: Multinominal logistic model with networth as predictor

Variables	Fml	Infm	Both	Fml	Infm	Both	Fml	Infm	Both
		BchMk			Urb			Rrl	
Gender	0.83***	1.11***	0.97***	0.80***	1.20***	0.99***	1.15***	0.92***	0.84***
Marital Status	1.98***	1.14***	1.97***	1.86***	1.11***	2.32***	2.37***	1.15***	1.76***
Age	0.95***	0.98***	0.95***	0.94***	0.98***	0.94***	0.97***	0.97***	0.96***
Employed	1.60***	1.06***	1.56***	1.61***	1.15***	1.54***	1.10***	0.87***	1.12***
Education	2.12***	0.62***	1.19***	2.22***	0.57***	1.30***	1.34***	0.84***	0.91***
Party	1.45***	0.78***	1.14***	1.45***	0.73***	1.15***	1.29***	0.81***	1.20***
HR	1.56***	0.64***	0.74***	——	——	——	——	——	——
Region-East	0.79***	0.66***	0.58***	0.86***	0.64***	0.58***	0.65***	0.61***	0.58***
Region-Center	0.64***	1.03***	0.76***	0.70***	1.07***	0.75***	0.49***	0.97***	0.79***
Fin.Inter	1.24***	0.98***	1.12***	1.24***	0.90***	1.00***	1.23***	0.99***	1.12***
Fin.Knowledge	2.14***	0.79***	2.19***	2.06***	0.78***	1.85***	2.06***	1.03***	3.22***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
NW-HE	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1	1.00***
Observations		26212			17393			8818	
AIC		42,422.85			27,303.32			14,986.46	
		Educ.0			Educ.1			CCP.0	
Gender	0.94***	1.12***	0.96***	0.81***	1.05***	0.84***	0.86***	1.17***	0.97***
Marital Status	1.83***	1.04***	1.51***	2.12***	1.12***	2.92***	1.48***	1.15***	2.57***
Age	0.95***	0.97***	0.95***	0.94***	0.99***	0.95***	0.94***	0.97***	0.94***
Employed	1.14***	0.99***	1.41***	1.94***	1.29***	1.93***	1.69***	1.20***	2.27***
Education	——	——	——	——	——	——	2.02***	0.73***	1.31***
Party	1.33***	0.77***	1.07***	1.60***	0.77***	1.26***	——	——	——
HR	1.35***	0.66***	0.69***	2.68***	0.42***	0.91***	1.83***	0.50***	0.60***
Region-East	0.76***	0.61***	0.57***	0.86***	0.81***	0.57***	1.00***	0.54***	0.61***
Region-Center	0.61***	1.04***	0.82***	0.71***	1.17***	0.69***	0.90***	0.97***	0.85***
Fin.Inter	1.24***	1.02***	1.12***	1.20***	0.72***	0.92***	1.11***	0.77***	1.18***
Fin.Knowledge	2.47***	0.82***	3.26***	1.90***	0.86***	1.95***	1.60***	0.57***	2.05***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00**	1.00***
NW-HE	1.00***	1.00***	1.00***	1.00***	1.00***	1.00*	1.00***	1.00*	1.00**
Observations		17804			8407			21183	
AIC		27,899.41			14,372.66			7,732.85	
		CCP.1			Sex.0			Sex.1	
Gender	0.84***	1.15***	0.98***	——	——	——	——	——	——
Marital Status	1.98***	1.13***	1.97***	1.63***	1.00***	1.66***	2.11***	1.22***	2.91***
Age	0.95***	0.98***	0.95***	0.94***	0.97***	0.96***	0.95***	0.98***	0.94***
Employed	1.57***	1.00***	1.36***	1.36***	1.03***	1.36***	1.80***	1.09***	1.45***
Education	2.11***	0.64***	1.10***	2.21***	0.57***	1.41***	2.14***	0.61***	1.16***
Party	——	——	——	1.45***	0.86***	1.08***	1.49***	0.78***	1.15***
HR	1.54***	0.66***	0.81***	2.05***	0.52***	0.66***	1.62***	0.69***	0.76***
Region-East	0.79***	0.65***	0.55***	0.74***	0.71***	0.57***	0.78***	0.60***	0.59***
Region-Center	0.59***	1.01***	0.74***	0.60***	1.12***	0.80***	0.63***	1.02***	0.79***
Fin.Inter	1.25***	1.02***	1.09***	1.34***	0.97***	1.03***	1.20***	0.96***	1.09***
Fin.Knowledge	2.21***	0.98***	2.39***	1.74***	0.66***	1.78***	1.94***	0.93***	2.16***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00**	1.00***	1.00***	1.00***
NW-HE	1.00***	1.00***	1.00***	1.00***	1.00***	1	1.00***	1.00***	1.00***
Observations		5028			6219			19992	
AIC		34,802.33			9,505.20			32,826.37	

Note: Multinominal Logistic Model with NW-HE (Net-Worth minus Home Equity) as Predictor. Relative Risk Ratios (RRR) are reported with significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. The number of the observation is approximately 80% of the total observation and the total number of observation may be vary based on the data-split group. The variables and abbreviations can be found in Appendix B.1 and Appendix B.2, respectively.

Table 3.10: MLM.2: Multinominal logistic model with NW-HE (Net-Worth minus home equity) as predictor.

Variables	Fml	Infm	Both	Fml	Infm	Both	Fml	Infm	Both
		BchMk			Urb			Rrl	
Gender	0.84***	1.11***	0.97***	0.81***	1.20***	1.00***	1.16***	0.92***	0.84***
Marital Status	1.93***	1.15***	1.92***	1.82***	1.11***	2.26***	2.29***	1.15***	1.71***
Age	0.95***	0.98***	0.95***	0.94***	0.98***	0.94***	0.97***	0.97***	0.96***
Employed	1.61***	1.07***	1.57***	1.62***	1.15***	1.55***	1.12***	0.87***	1.13***
Education	2.07***	0.62***	1.16***	2.16***	0.57***	1.26***	1.32***	0.84***	0.89***
Party	1.43***	0.78***	1.13***	1.44***	0.73***	1.13***	1.26***	0.80***	1.18***
HR	1.51***	0.64***	0.72***	——	——	——	——	——	——
Region-East	0.77***	0.66***	0.56***	0.82***	0.64***	0.55***	0.63***	0.61***	0.56***
Region-Center	0.64***	1.03***	0.76***	0.70***	1.07***	0.75***	0.49***	0.97***	0.79***
Fin.Inter	1.22***	0.98***	1.10***	1.22***	0.91***	0.98***	1.23***	0.99***	1.12***
Fin.Knowldge	2.07***	0.79***	2.10***	1.98***	0.78***	1.77***	1.96***	1.02***	3.07***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
Liquid Assets	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1	1.00***
Observations		26212			17393			8818	
AIC		42,258.51			27,147.32			14,944.02	
		Educ.0			Educ.1			CCP.0	
Gender	0.95***	1.12***	0.97***	0.81***	1.05***	0.84***	0.87***	1.17***	0.98***
Marital Status	1.79***	1.04***	1.48***	2.06***	1.13***	2.82***	1.44***	1.15***	2.50***
Age	0.95***	0.97***	0.95***	0.94***	0.99***	0.95***	0.94***	0.97***	0.94***
Employed	1.15***	0.99***	1.42***	1.95***	1.29***	1.94***	1.72***	1.19***	2.31***
Education	——	——	——	——	——	——	1.96***	0.73***	1.28***
Party	1.33***	0.77***	1.06***	1.57***	0.77***	1.23***	——	——	——
HR	1.31***	0.66***	0.67***	2.56***	0.43***	0.86***	1.76***	0.51***	0.58***
Region-East	0.74***	0.61***	0.56***	0.81***	0.81***	0.53***	0.94***	0.54***	0.57***
Region-Center	0.61***	1.04***	0.83***	0.71***	1.16***	0.69***	0.90***	0.97***	0.84***
Fin.Inter	1.23***	1.02***	1.11***	1.18***	0.72***	0.90***	1.10***	0.77***	1.17***
Fin.Knowldge	2.39***	0.82***	3.16***	1.83***	0.86***	1.86***	1.57***	0.57***	1.99***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00*	1.00***
Liquid Assets	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00**	1.00***
Observations		17804			8407			21183	
AIC		27,825.96			14,286.07			7,681.80	
		CCP.1			Sex.0			Sex.1	
Gender	0.85***	1.15***	0.98***	——	——	——	——	——	——
Marital Status	1.93***	1.13***	1.93***	1.59***	1.00***	1.62***	2.06***	1.23***	2.84***
Age	0.95***	0.98***	0.95***	0.94***	0.97***	0.96***	0.95***	0.98***	0.94***
Employed	1.58***	1.00***	1.37***	1.38***	1.03***	1.38***	1.80***	1.09***	1.46***
Education	2.07***	0.65***	1.07***	2.16***	0.57***	1.37***	2.08***	0.62***	1.13***
Party	——	——	——	1.42***	0.86***	1.06***	1.48***	0.78***	1.14***
HR	1.49***	0.66***	0.78***	1.99***	0.52***	0.64***	1.57***	0.69***	0.74***
Region-East	0.77***	0.65***	0.53***	0.70***	0.71***	0.53***	0.76***	0.60***	0.57***
Region-Center	0.60***	1.01***	0.75***	0.60***	1.12***	0.80***	0.64***	1.01***	0.80***
Fin.Inter	1.23***	1.02***	1.08***	1.32***	0.97***	1.00***	1.18***	0.96***	1.07***
Fin.Knowldge	2.14***	0.99***	2.30***	1.71***	0.66***	1.74***	1.87***	0.94***	2.07***
Income	1.00***	1.00***	1.00***	1.00***	1.00***	1	1.00***	1.00***	1.00***
Liquid Assets	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
Observations		5028			6219			19992	
AIC		34,677.61			9,458.10			32,703.07	

Note: Multinominal Logistic Model with Liquid Assets) as Predictor. Relative Risk Ratios (RRR) are reported with significance: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. The number of the observation is approximately 80% of the total observation and the total number of observation may be vary based on the data-split group. The variables and abbreviations can be found in Appendix B.1 and Appendix B.2, respectively.

Table 3.11: MLM.3: Multinominal logistic model with liquid assets as predictor.

Data splits	$y = Access\,Loan$			$y = Loan\,Type$		
	1	2	3	1	2	3
	GLM			MLM		
Urb & Rrl	0.699	0.696	0.699	0.710	0.707	0.710
Educ.0 & Educ.1	0.685	0.683	0.686	0.712	0.708	0.712
CCP.0 & CCP.1	0.708	0.706	0.708	0.720	0.717	0.720
SEX.0 & SEX.1	0.680	0.677	0.681	0.705	0.702	0.706
BchMk	0.698	0.695	0.699	0.712	0.709	0.709
	BAG			BAG		
Urb & Rrl	0.668	0.661	0.663	0.664	0.658	0.659
Educ.0 & Educ.1	0.662	0.655	0.644	0.671	0.676	0.668
CCP.0 & CCP.1	0.664	0.667	0.662	0.676	0.676	0.676
SEX.0 & SEX.1	0.659	0.662	0.653	0.671	0.666	0.676
BchMk	0.667	0.664	0.660	0.669	0.677	0.666
	RF			RF		
Urb & Rrl	0.688	0.687	0.685	0.677	0.674	0.680
Educ.0 & Educ.1	0.672	0.669	0.668	0.679	0.678	0.680
CCP.0 & CCP.1	0.691	0.690	0.690	0.686	0.685	0.690
SEX.0 & SEX.1	0.670	0.672	0.664	0.683	0.675	0.679
BchMk	0.687	0.683	0.682	0.689	0.682	0.687
	GBM			GBM		
Urb & Rrl	0.718	0.722	0.716	0.721	0.719	0.722
Educ.0 & Educ.1	0.700	0.701	0.700	0.726	0.722	0.726
CCP.0 & CCP.1	0.721	0.725	0.718	0.732	0.733	0.733
SEX.0 & SEX.1	0.699	0.700	0.694	0.722	0.720	0.724
BchMk	0.717	0.718	0.722	0.725	0.725	0.729

Note: AUC of the Model (1:3) are shown. Where Model (1:3) presents Networth, NW-HE, and Liquid Assets, as predictor, respectively. The abbreviations can be found in Appendix B.2.

Table 3.12: AUC of all the models

Chapter 4

Predicting Financial Health of the Households Using Machine Learning Algorithms

4.1 Introduction

After the Communist Party won the civil war which lasts more than 20 years, private enterprises were fully banned in China between 1952 and 1977. The new economic reforms legislated in 1978, policies target the urban sector (Wan, 2008). Most of the major formal financial service providers (e.g., state-owned and commercial banks as well as rural credit cooperatives), providing external finance to a vast number of Chinese households, living in rural or poor areas (Hannig and Jansen, 2010), were closed (Sparreboom and Duflos, 2012). Meanwhile the urban-rural gap was widening by the excessive manufacturing production.

Subsequently, income inequity has increased tremendously. Study of Wan and Sebastian (2011) reported that, in 2008, 336 million Chinese living on under $2.00 a day (purchasing power parity or PPP-adjusted), more than 100 million Chinese surviving on no more than $1.25/day (PPP-adjusted). On the contrary, in 2010, the People's Republic of China (PRC) had 960,000 millionaires, each with more than $1.6 million in personal wealth, and in the following year in the PRC there were 146 billionaires, each with more than $1 billion in assets. Within this context, reducing income inequity and promoting access to finance has become a policy concern.

One of the major reasons for poor remaining poor is linked to lack of access to formal credit (Collins et al., 2009) and the existence of a large gap between demand and supply (Sparreboom and Duflos, 2012). Understating the barriers for accessing finance is crucial, barriers to access to finance can be either geographic (e.g., absence of nearby bank branches) or socioeconomic (e.g., minimum income & high collateral requirements, social, or ethnic groups) (Hannig and Jansen, 2010). Indeed, over the last decade, the Chinese government has implemented a number of new policies to promote financial inclusion; these implementations ground a great number of bank outlets

and formal financial service points (Sparreboom and Duflos, 2012). In spite of recent policy efforts, the study of the World Bank showed that the usage of formal credit in China remains still scarce comparing high-income economies (Demirguc-Kunt and Klapper, 2012). Its credit market remains undeveloped (Sparreboom and Duflos, 2012), and the disparity of accessing social services (e.g., health care, education, etc.) has continued to grow (Li et al., 2008; Wang et al., 2014; Zhuang, 2008).

In order to assess the limitations on accessing credit, it is essential to analyze the use of financial services. However, one of the most reliable and up-to-date data source, Global Findex, analyses bank-level data which mostly consider state-owned firms, and it has so limited observations since its main purpose is to compare financial inclusion across 148 countries (Fungáčová et al., 2014). A rising literature has analyzed the characteristic of households to access to finance, the probability of default, and preventing the discrimination, but a few have focused on the household's financial strength and spatial distribution of poor households.

Thereby, the aim of this study is to propose a predictive model of household financial conditions using machine learning. From a policy perspective, this tool may help policy makers to localize the disadvantaged Chinese households to promote financial inclusion by expanding access to socially and financially disadvantaged groups. The statistical analysis is based on individual-level data from the 2015 China Household Financial Survey (CHFS), which focuses on a sample of 133,183 individuals, including 37,289 household heads. First, we employed unsupervised learning (K-means clustering) to classify Chinese households's financial strength (namely, Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, and Cluster 6, where Cluster 1 and Cluster 4 are less populated wealth households, Cluster 2 and Cluster 6 are highly populated poor households, and Cluster 3 and Cluster 5 have slighly moderate wealth).

Secondly, we employed supervised learning (bagging (BAG), random forest (RF), and boosting (GBM)) for prediction. There is a wide literature on access to finance based on the applicant characteristics using classical methods. However, to our best knowledge there has been little research done to classify household financial health, using unsupervised and supervised learning approaches.

We choose China for mainly four reasons. Firstly, China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018). Secondly, China is the world's first-largest economy and has rapid economic growth over the past few decades. Yet, China is

still considered a developing country, and millions are below the international poverty standards. Finally, the use of formal financial services (e.g., bank loans and credit cards) is much lower than other emerging economies (e.g., Brazil, India, Russia, and South Africa) (Fungáčová et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012). Thirdly, in 2005 self-employment accounted for 10% of the urban labor force, which means that at least 64 million people were self-employed in China with 40 million were rural-urban migrants and 24 million were self-employed urban residents (Cui et al., 2013). Only in 2015, China had generated 117 million self-employed jobs.

Our study contributes to policy making in a number of ways. First, clustering Chinese households based on their financial strength can help policy makers to identify and target specific economically disadvantaged Chinese households. Secondly, We can see the proportion of access to finance for each cluster. And lastly, we built a Shinyapp[1], using CHFS 2015 to guide policy makers.

This study focus on predictive rather than exploratory modeling and the rest of the paper is structured as follows. Section 4.2 presents the related literature review. Section 4.3 presents the data description. Section 4.4 describes the methodology. Section 4.5 discusses the results of this study. Section 4.6 introduces the Shinyapp.

4.2 Literature Review

Literature has long debated on how to effectively measure household finance and financial inclusion. Demirguc-Kunt and Klapper (2013) argue that financial inclusion can be measured based on the three main indicators: 1) *Formal account:* having an account, 2) *Formal saving:* saving behavior, and 3) *Formal credit:* the usage of bank credit in a formal financial institution.

Analyzing the Global Findex data, the study of Fungáčová and Weill (2015) show that 66% of Chinese have a formal account; Thus, half of the world population does not have a formal account (Demirguc-Kunt and Klapper, 2013). In the past year, 82% of Chinese individuals have saved at a formal financial institution; Whereas world average formal saving is 22% (Demirguc-Kunt and Klapper, 2013). Notwithstanding, in the past 12 months, not more than 7% of Chinese individuals have used formal credit, which remains highly concentrated in large state-owned Chinese firms

---

[1]See at: https://seymakalay87.shinyapps.io/micro/

(Hale and Long, 2011).

A common tool to assess individuals'or households'financial strength is the use of credit scores. Credit scoring is a way of differentiation between two type of loan applicants (good and bad). When the loan applicant will most likely pay back his/her debt on time, it is called good otherwise bad loan (or credit). The earliest studies on credit scoring can be traced back to Fisher (1936), and the same idea was firstly implemented by Durand (1941) to define good and bad credits. Up to now, many traditional statistical methods and advanced machine learning (ML) algorithms are implemented to determine the creditworthiness of applicants. Great among of studies have focused on improving the model prediction power. Indeed, these highly accurate machine learning methods (such as artificial neural networks (ANNs: Desai et al. (1996); West (2000)), decision trees (DTs: Hung and Chen (2009)), support vector machines (SVMs: Huang et al. (2007); Schebesch and Stecking (2005), etc.)), playing a crucial role in our everyday lives, are intend to help human in various decision-making concepts (Montgomery et al., 2000; Mysiak et al., 2005; Yoshimura et al., 2006).

Although, ML algorithms are considered to be powerful prediction methods, typically, they are considered being as a black-box and their results can be biased since the historical data sample is biased and it can exclude financially disadvantaged groups systematically in terms of access to finance. As ONeil (2017) emphasizes we can create great social benefits as well as we can bring devastation using machine learning.

Clustering techniques may be used to define similar subgroups within a data set. The observations in the same subgroups are alike, and in the different subgroups are unlike. Clusters have numerous advantages in the economy for firms, institutions, and individuals. The study of Wolman and Hincapie (2010) present various conceptualizations of clustering. Further, mapping clusters can help to raise awareness, to design suitable programs by regions, to fight poverty, and to promote financial inclusion.

Henninger and Snel (2002) reported several poverty maps based on consumption expenditures: from national to regional, provincial, and municipal levels. They found that poor people are geographically spread around the country, yet they tend to be clustered in specific places.

For example, In Ecuador, the lowest-poverty region contains provinces and municipalities with poverty rates range from 45% to 75%. In 2000, 112 geographic areas of 62 provinces were mapped

in Vietnam. The analysis shows that minority groups are mostly located in poorer areas. In 2001 in Guatemalan, poverty and available roads were mapped, showing that poverty is extremely high in those regions where there are non-asphalt roads. In South Africa, poverty and the spread of cholera was mapped, showing that initial disease in 2001 started from the poorest regions and spread around to other poor and highly poor regions.

To enable multidimensional poverty patterns, in 2014 Källestål et al. (2020) clustered Unsatisfied Basic Needs (UBN) index, using *K*-means and SimpleKMeans clustering algorithms, on Cuatro Santos Health and Demographic Surveillance survey (HDSS), including 5.966 household level observations. They defined six clusters namely *poorest*, *poor*, *fairly poor*, *fairly rich*, *rich*, and *richest*. They found around 64% of the households, who are in the poor cluster, reported various degrees of food insecurity, whereas rich households never reported it. Households, in the poor cluster, were still using horses and tortilla oven, while, in rich clusters, households were using motorbikes and have computers. In addition, the study showed that, in the rich cluster, the proportion of female household heads were higher.

Using the health and demographic survey system (HDSS) data-set, McParland et al. (2014) clustered approximately 82.000 households based on their socio-economic status, and they found the mortality rates differ based on the socio-economic groups. Using principal component analysis (PCA), Amarasinghe et al. (2005) mapped the poverty in Sri Lanka. They identified two clusters, and in the high-level poverty cluster, the dominant economic activity is agriculture, and better access to roads is unlikely compared to other clusters.

Ali et al. (2014) examined around 17,000 micro-enterprises in four regions of Ethiopia and four sectors. The cottage/handicraft manufacturing survey was obtained from 2002/2003 and conducted by the Central Statistical Agency of Ethiopia (CSAE). They found that industrial clusters create iterative interactions between traders and local producers that generates trust that may reduce the moral hazard by enabling a reciprocal exchange of information. The study of UNCTAD (2005) mentioned that the clustering strategy can help to promote SMEs financially and non-financially.

Major part of the Chinese poor households, inhabits rural areas (Gale et al., 2005), can not full fill the socioeconomic requirements to enter the formal loan. Thus, without finance, they can not take advantage of opportunities such as investment, education, and health care. Most of the literature also points out the role played by microfinance in favoring equal opportunities for the

poor in developing countries, including investments in capital assets, education, and healthcare. Indeed, Yunus (1998), founder of the Grameen Bank[2], addressed in his book the importance of micro-finance and the lack of credit for the skillful entrepreneurs in impoverished countries. He discusses how the lives of the borrowers, businesses, and local areas can be changed by giving an opportunity to access micro finance. For example, using the household survey conducted by the Bangladesh Institute for Development Studies (BIDS) in 1998-99 including 1,798 households from 87 villages, Pitt et al. (2006) show that finance provided to women (who are more credit constrained than men) improves the health and nutrition of both boys and girls. The study shows that women'empowerment increases together with participating in microfinance programs.

In the mid-1990s, as part of the Chinese government's poverty reduction strategies, micro-finance was introduced. In recent years, outstanding agricultural loans were more than double between 2001-2005 and reached \$127 billion in 2005 (Gale and Callender, 2006). However, in China, Li and Zhao (2011) showed that most of the clients of the largest micro credit provider, Rural Credit Cooperative (RCC), are relatively wealthy since micro-credit institutions are not suitable for the poor and they are not targeting the poorest households.

4.3 Data Description

The statistical analysis is based on the 2015 China Household Finance Survey (CHFS), which was nationally implemented by The Survey and Research Center for China Household Finance, established in 2010 as a non-profit research institution. CHFS documents the micro-level household finance from all over China, which includes socioeconomic (e.g., savings, cash, etc.) and geographic (e.g., region, household registration, province, etc.) characteristics of the households.

The third wave of the CHFS household-level survey was implemented with a three-stage probability proportion to size (PPS) sampling performed. In the first stage of sampling, 29 provinces (except Hong Kong, Inner Mongolia, Macau, Tibet, Taiwan, and Xinjiang), 351 counties were selected. The second stage involves selecting committees/villages from the counties/cities chosen in the previous stage, and the last stage involves selecting households from the 1396 committees/villages selected in the earlier stage with a total sample size 37,289. In addition to that small

---

[2]The Grameen Bank, founded by Yunus in 1977, won the Nobel Peace Prize in 2006 for its positive impact on helping poverty.

and micro enterprises covered 28 provinces, 79 counties, and 234 towns with the sample size was 5497 households.

4.3.1 Dependent Variables

In the 2015 CHFS survey, participants were ask if they have any formal loan and/or if they engage in any other loan besides formal loan, the column wise percentage for the purpose of the loan (Table 3.1) as well as the reasons for not having formal loan in Table 3.2.

In the third wave of the CHFS survey, the participants were asked the reason for not having loan. Table 3.2 shows that 37% of the households were reported that they do not need loan. 45% never applied loan and 37.3% of them believe the loan would not be approved. The main reasons of the loan rejection were with 28.85% not having guarantor, with 29.69% having low income, with 24.37% not being familiar with the loan officer, or with 23.81% not having collateral while only 4.76% of them has recorded having a bad credit history.

Table 3.1 prints out the purpose of the loan and access to multiple loan is allowed, for example having housing and vehicle formal loan together. So the households' head obtain mainly formal loans from credit cards (45.78%), need informal loan for housing (39.73%) and medical(27.7%) purposes, and need both loan for housing (47.39%) and micro-enterprise(47.39%).

The dependent variables were assigned as follows, $access\,to\,credit = 1$ if respondent has at least one of the loans (e.g., formal, informal, or both) for any purposes, otherwise $access\,to\,credit = 0$.

$Formal = 1$ if the participant has only borrowed from formal institutions (e.g., banks) or had had credit card debt, $Informal = 1$ if the participant has only borrowed from informal organizations (e.g., money lenders) or network (e.g., family, friends), and $Both = 1$ if the participant has both formal and informal loans, and $No\,Loans = 1$ if the participant has non of the mentioned loans for any kind of purposes[3].

4.3.2 Explanatory Variables

Variables were set up as follows gender (male = 1, female = 0), married (married or cohabited = 1, otherwise = 0), employed (employed = 1, otherwise = 0), education (high school or higher = 1, otherwise = 0), party (affiliation with Chinese communist party = 1, otherwise = 0),

---

[3]Note: If participant has no loan $No\,Loans = 1$ indicates that he/she has not access to loan $access\,to\,credit = 0$

household registration (HR) (urban = 1, rural = 0), region (west, east, and center), Fin.Knowledge (Fin.Knowledge = 1 when household's head took ant financial class or had defined him/herself has having financial knowledge well, otherwise = 0), Fin.Inter (Fin.Inter = 1 when household's head is located 1km arrange of a formal institution, otherwise = 0) and age, income, and 3 asset holding variable were coded as continues. We define the household asset holding assets, based on the previous studies (Brandolini et al., 2010; Caner and Wolff, 2004; Haveman and Wolff, 2004; Huang et al., 2013), and we measure households'asset holdings with three different continuous variables: net worth (the value of financial and non-financial assets minus liabilities), net worth minus home equity (NW-HE), and liquid assets (cash and other easily cash-able assets). Definition of the variables and abbreviations can be found Appendix B.1 and Appendix B.2, respectively.

4.3.3 Descriptive Statistics

Table 4.1 shows the summary statistics of the Chinese Households as a whole, after cleaning the CHFS survey. 76% of the respondents were male headed households and almost 86% of the households' head were married with an average age 54. The majority of the households (48%) are prefer to live in the east. 19% of the households have a political affiliation with the current Chinese Communist Party. Only 32 % of the households had a high school or higher education. The average annual income was 69731.55 CNY (SD = CNY 175283.50) and our three asset owning variables were networth, NW-HE and liquid assets were CNY 745862.37 (SD = CNY 1603699.60), CNY 708689.99 (SD = CNY 1526658.33), 799474.99 (SD = CNY 1633306.20), respectively. Our primary analysis shows that having access to formal, informal, both (formal and informal) and having not access to any type of loan are 11.65%, 13.96, 4.79%, and 69.58%, respectively.

4.4 Methodology

Access to finance may depend on the experimental and judgmental heuristic matters. Such an approach can be highly subjective, and robust. Another words, unsupervised learning algorithm help us to identify subgroups in the data-set based on the observations' similarities which results a subjective decision making about access to finance. One of the most commonly used unsupervised ML technique is K-means clustering, and it can guarantee the homogeneity within the same clusters

Variable	Factor						
	0			1			
Gender	Female	23.72%	Male	76.27%			
Married	Others	13.99%	Married/Cohabited	86.00%			
Employed	Others	35.89%	Employed	64.10%			
Education	Others	67.92%	High-school or above	32.07%			
Party	Others	80.81%	Party	19.18%			
HR	Rural	33.64%	Urban	66.35%			
Fin.Knowledge	Others	94.54%	Fin.Knowledge	5.45%			
Fin.Inter	Others	81.53%	Fin.Inter	18.46%			
Access Loan	Others...........................	69.58%	Access Loan	30.41%			
Formal	Others	88.35%	Formal	11.65%			
Informal	Others	86.04%	Informal	13.96%			
Both	Others	95.21%	Both	4.79%			
No Loans	Others	30.41%	No loans	69.58%			

	West	East	Center
Region	27.25%	48.26%	24.48%

	Continuous						
	Mean	SD	Median	Min	Max	Skew	Kurt
Age	53.90	14.28	53.0	17.0	101	0.02	-0.50
Income	69731.55	175283.50	41750.0	-800000.0	5000000	16.70	396.64
Networth	745862.37	1603699.60	283353.3	-627904.2	19999748	6.50	58.35
NW-HE	708689.99	1526658.33	270636.5	-3614776.4	19999748	6.53	59.59
Liquid Assets	799474.99	1633306.20	325099.6	0.0	20000000	6.45	57.53

Note: HR stands for Household Registration. NW-HE is net-worth minus home equity. All the asset variables (e.g. income, net-worth, NW-HE, and liquid assets are in Chinese renminbi (CNY).

Table 4.1: Summary statistics

(e.g., applicants, households, etc.) with a low variance in a cluster (Hartigan and Wong, 1979). After grouping the data-set with K-means clustering based on the households' financial strength, we applied supervised learning (e.g, classification trees) to predict the right cluster of the households. A classification tree is used to predict a qualitative response such as categories/class predictions (Kuhn, 2008).

Following James et al. (2013), we used 10 cross-validation (cv) based on 80/20 train/test sets. We implemented stratified random sampling on our dependent variables for both access to loan and its type then, we divided the data-set into 80% where we set our models and we compare the prediction results with the 20% of the data-set and to make sure we are not over-fitting we did this 10 times (10 different train and test sets), then we took the average of the model performances.

	Income	Networth	Liquid Assets	NW-HE
Income	1***			
Networth	0.45***	1***		
NW-HE	0.43***	0.98***	0.97***	1***
Liquid Assets	0.46***	0.96***	1***	

Table 4.2: Correlation Matrix

Table 4.2 shows the correlation matrix of the continuous (income and our three assets) variables, I remove highly correlated networth, and I clustered NW-HE, and liquid assets.

4.4.1 Unsupervised Learning: K-Means Clustering

k-means clustering algorithm starts by randomly selecting k objects (also known as centroids) from the data-set as the initial clusters' center. Then "cluster assignment step" takes place by minimizing the Euclidean distance (Eq. 4.2) between the remaining object (k) and its centroid. The algorithm calculates the final mean value for each cluster. After the "centroid update step" all the objects are reassigned based on the updated cluster means. These steps iteratively repeated until convergence is achieved (another word until the cluster assignments stop changing)[4]. K-means clustering algorithm is "the most popular and the simplest partitional algorithm" (Jain, 2010).

Let C_k be the generic cluster of observations with $k = 1, .., K$ which satisfies the two properties:

1. $C_1 \cup C_2 \cup ... \cup C_K = \{1, ..., n\}$, where each observation belongs to at least one cluster.

---

[4]k is the pre-specified number of clusters.

2. $C_k \cap C_{k'} = 0$ for all $k \neq k'$. Each observation belongs only one cluster.

In other words, i-th observation belongs to k-th cluster only, then $i \in C_k$. K-means cluster sets the *within − cluster variance* as small as possible, and K-means cluster is defined as

$$\min_{C_1,\ldots,C_K} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\} \tag{4.1}$$

where $|C_k|$ is the number of observation in the k-th cluster, p is the vector of futures, and $W(C_k)$ is a measure of distance, called *squared Euclidean distance* and it is defined as

$$W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$$

gives us the best clustering of the observations which minimize the *within − cluster variance* by minimizing the *squared Euclidean distance*. Algorithm 4 laid out the K-means optimization problem defined in Equation 4.1.

Algorithm 4 K-Means Clustering

For an initial clustering assign a random number (from 1 to K) to each observation.
Iterate the following steps until the cluster assignments stop changing.
  1: Compute the cluster centroid (center) for each cluster.
  2: Using the Euclidean distance, to assign each observation to the closest centroid.

4.4.2 Supervised Learning: Classification Trees

The aim of classification trees is to split the data into smaller and homogeneous groups. If the notes of the split are pure, it can be said groups are homogeneous (i.e., containing a huge proportion of a single class in each node). Classification trees are used to build and evaluate models for a categorical response and the performance of the classification trees can be measure by Gini index.

Gini index (Breiman et al., 1984) is a measure of total heterogeneity across the K classes and it is minimized when class probabilities $(\hat{\pi}_{mk})$ is driven towards zero or one. Gini index is a purity measure which tries to maximize accuracy or minimize the misclassification error and it is defined as

$$G_m = \sum_{k=1}^{K} \hat{\pi}_{mk}(1 - \hat{\pi}_{mk}) \tag{4.2}$$

where $0 \leq \hat{\pi}_{mk} \leq 1$ and $\hat{\pi}_{mk}$ is the portion of training observations in the m-th note that belong to the k-th category/class.

4.4.2.1 Bagging

Decision trees suffer from high variance[5]. Bagging (shown in Algorithm 5) is an ensemble procedure which reduces the variance and increases the prediction accuracy of a statistical learning method by considering many training sets $(\hat{f}^1(x), \hat{f}^2(x), ..., \hat{f}^B(x))$ from the population. Since we can not have multiple training-sets, from a single training data-set, we can generate B different bootstrapped training data-sets $(\hat{f}^{*1}(x), \hat{f}^{*2}(x), ..., \hat{f}^{*B}(x))$ by each B trees and take a majority vote. Therefore, bagging for classification problem defined as

$$\hat{f}(x) = \operatorname*{argmax}_{k} \hat{f}^{*b}(x) \tag{4.3}$$

Algorithm 5 Bagging

1: **for** $i = 1$ to m predictions **do**
2:     From the original data-set, generate a bootstrap sample.
3:     Train an unpruned classification tree on this sample.
4: **end for**

4.4.2.2 Random Forest

Rather than considering the random sample of m predictors from the total of p predictors in each split, random forest does not consider a majority of the p predictors, and considers in each split a fresh sample of m_{try} which we usually set to $m_{try} \approx \sqrt{p}$ (Breiman et al., 1984).

Random forests which de-correlate the trees by considering $m_{try} \approx \sqrt{p}$ show an improvement over bagged trees $m = p$. Random forest can be constructed as shown in Algorithm 6 (Breiman et al., 1984).

---

[5]If we split the training data-set randomly into two parts and set a decision tree to both parts, the results might be quite different.

Algorithm 6 Random Forest

1: Select m, the number of models to build.
2: **for** $i = 1$ to m **do**
3:     Generate a bootstrap sample of the original data.
4:     Based on this sample, train a classification tree model.
5:     **for** each split **do**
6:         Randomly select k predictors out of P of the original predictors, where k $(< P)$.
7:         Select the best predictor among the k predictors and partition the data
8:     **end for**
9:     Use typical tree model stopping criteria to determine when a tree is complete.
10: **end for**

4.4.2.3   Gradient Boosting

Unlike bagging trees, boosting does not use bootstrap sampling, rather each tree is fit using information from previous trees. An event probability of stochastic gradient boosting model (shown in Algorithm 7) is given by

$$\hat{\pi}_i = \frac{1}{1 + exp[-f(x)]'} \tag{4.4}$$

where $f(x)$ is in the range of $[-\infty, \infty]$ and its initial estimate of the model is $f_i^{(0)} = log(\frac{\pi_i}{1-\pi_i})$, where $\hat{\pi}$ is the estimated sample proportion of a single class from the training set.

Algorithm 7 Gradient Boosting

1: Initialized all predictions to the sample log-odds: $f_i^{(0)} = log(\frac{\pi_i}{1-\pi_i})$
2: **for** *iteration* $j = 1...M$ **do**
3:     Compute the residual $z_i = y_i - \pi_i$.
4:     Randomly sample the training data-set.
5:     Train a tree model.
6:     Compute the terminal node estimates of the residuals.
7:     $r_i = \frac{\frac{1}{n}\sum_i^n (y_i - \pi_i)}{\frac{1}{n}\sum_i^n \pi_i(y_i - \pi_i)}$.
8:     Update the current model $f_i = f_i + \lambda f_i^{(0)}$, where λ is a shrinkage parameter.
9: **end for**

4.5 Results

4.5.1 Unsupervised Learning: K-Means Clustering

In this section, we analyze the results of the K-means clustering. Using K-means, we clustered NW-HE then, we set $K = 6$ clusters based on the model performance in Table 4.3. We can see when the cluster number $K = 6$ models predictive powers are (MLM = 71.30, BAG = 68,94, and RF = 61.52) are higher comparing to others. To avoid any confusion, we will only interpret MLM.

We can see from Table 4.4, the number of households[6]. The household's population ranges between clusters from 117, Cluster 1 the less populated, to 22275, Cluster 6 the most populated. Between clusters the age variable shows similarity, but income differs. Mean income in Cluster 6 41603.10 CNY and in Cluster 2 is 85273.24 seems to most populated and less wealthy households comparing to other clusters. Cluster 1, with population 117, and Cluster 4, with population 284, have the first and second highest average mean income. Additionally, we present descriptive statistics for factor (Table 4.5 and Table 4.6) variables.

	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$	$K = 8$	$K = 9$	$K = 10$
MLM	69.00	65.37	70.68	71.30	62.32	62.28	56.70	53.25
BAG	49.92	57.24	63.26	68.94	49.62	54.22	57.62	56.66
RF	43.03	64.16	55.23	61.52	53.39	63.98	49.97	55.41
GBM	-	-	-	-	-	-	-	-

Note: Results were GBM were not converged for all the clusters, so we drop it.

Table 4.3: AUC of each model by cluster

Table 4.5 shows us that relatively wealthy households (Cluster 1, 3, 4, and 5) are located in East. Cluster 6 is almost equally distributed among the HR but relatively higher in rural area with 56% of the observations. As we have noted above, urban targeting reforms in 1978 might cause East to be highly industrialized area, mostly populated by middle and upper income groups.

Table 4.6 can be interpenetrated as follow: Cluster 6, the poorest and most populated cluster, shows the lowest education level (22.41%), party affiliation (15.08%), financial knowledge (2.83%), access to formal loan (5.79%). Cluster 1, the richest and less populated cluster, shows a higher education level (63.25%), party affiliation (27.35%), financial knowledge (29.91%), and the highest access to formal loan (51.28%) comparing to other clusters.

---

[6]Please note: Relatively wealthy regions and urban areas were over-sampled in CHFS survey.

	Mean	Sd	Median	Min	Max	Skew	Kurt
	Cluster 1 (# 117 Households)						
Age	45.42	12.5	46	19	77	0.09	-0.27
Income	1.092860e+06	1.343100e+06	500940	-124250	5000000	1.59	1.89
	Cluster 2 (# 6875 Households)						
Age	52.57	14.65	52.0	17.0	95	0.11	-0.59
Income	85273.24	110173.601	66694.0	-749573.0	2920000	8.98	162.23
	Cluster 3 (# 2387 Households)						
Age	52.931	14.98	52	17	93	0.13	-0.70
Income	1.462134e+05	277550.02	97810	-800000	5000000	11.29	173.63
	Cluster 4 (# 284 Households)						
Age	47.99	14.57	46	20	88	0.41	-0.41
Income	3.633347e+05	5.804826e+05	197250	-800000	5000000	4.28	28.17
	Cluster 5 (# 827 Households)						
Age	50.96	14.92	50	17	91	0.17	-0.71
Income	2.318359e+05	393834.50	143500	-800000	5000000	7.42	74.71
	Cluster 6 (# 22275 Households)						
Age	54.6	13.99001	54.0	17.0	101.0	-0.01	-0.41
Income	41603.10	68574.56425	30000.0	-338992.0	5000000.0	36.18	2479.99

Note: Kurt and Skew stand for kurtosis and skewness, respectively.

Table 4.4: Descriptive statistics of continuous variables by cluster

	Region			HR	
	East	Center	West	Urban	Rural
Cluster 1	74%	9%	17%	6%	94%
Cluster 2	57%	21%	22%	15%	85%
Cluster 3	76%	12%	12%	7%	93%
Cluster 4	82%	10%	8%	10%	90%
Cluster 5	83%	8%	9%	4%	96%
Cluster 6	41%	32%	28%	44%	56%

Note: HR stands for Household Registration (urban or rural).

Table 4.5: Distribution of regions and HR by cluster.

Additionally, Table 4.7 presents the Chi test, and we can conclude that access to loan and loan type is significantly different in each Cluster. Table 4.8 shows the proportion of households in the different clusters by province. We can see that in each province the population of poor (Cluster 2 and Cluster 6), is relatively high. And wealthier households (Cluster 1 and Cluster 4) are distributed almost equally around the regions.

The percentage calculation in Table 4.9 was done by cluster. Inequity in Guangdong seems to be the highest with the lowest financial health (7% of Cluster 6 and 8% of Cluster 2) and the

	Cluster 1		Cluster 2		Cluster 3	
	0	1	0	1	0	1
Gender	25.64%	74.36%	25.72%	74.28%	30.41%	69.59%
Married	15.38%	84.62%	11.13%	88.87%	11.48%	88.52%
Employed	23.08%	76.92%	38.85%	61.15%	41.64%	58.36%
Education	36.75%	63.25%	52.49%	47.51%	40.05%	59.95%
Party	72.65%	27.35%	74.4%	25.6%	68.79%	31.21%
Fin.Knowledge	70.09%	29.91%	91.51%	8.49%	86.89%	13.11%
Fin.Inter	82.05%	17.95%	78.34%	21.66%	78.3%	21.7%
Formal	48.72%	51.28%	80.33%	19.67%	72.22%	27.78%
Informal	98.29%	1.71%	92.64%	7.36%	95.43%	4.57%
Both	88.03%	11.97%	93.48%	6.52%	93.67%	6.33%
No.Loan	64.96%	35.04%	33.54%	66.46%	38.67%	61.33%
	Cluster 4		Cluster 5		Cluster 6	
	0	1	0	1	0	1
Gender	29.58%	70.42%	29.14%	70.86%	22.11%	77.89%
Married	8.45%	91.55%	9.92%	90.08%	15.36%	84.64%
Employed	33.1%	66.9%	37%	63%	34.42%	65.58%
Education	28.17%	71.83%	34.34%	65.66%	77.59%	22.41%
Party	67.25%	32.75%	63.97%	36.03%	84.92%	15.08%
Fin.Knowledge	78.52%	21.48%	80.41%	19.59%	97.17%	2.83%
Fin.Inter	76.06%	23.94%	77.75%	22.25%	83.08%	16.92%
Formal	54.93%	45.07%	60.58%	39.42%	94.21%	5.79%
Informal	97.89%	2.11%	96.74%	3.26%	82.37%	17.63%
Both	90.14%	9.86%	94.44%	5.56%	96.03%	3.97%
No.Loan	57.04%	42.96%	48.25%	51.75%	27.38%	72.62%

See at Section 4.3.2 for the definition of 0 and 1.

Table 4.6: Descriptive statistics of factors by cluster

	Access to Loan	Access to Loan Type
Chi-Test	0.011	0.000

Table 4.7: Chi Squared (χ^2) test was conducted based on the cluster.

highest financial health (31% of Cluster 1 and 23% of Cluster 4), comparing to other province.

4.5.2 Unsupervised Learning

In this section, we analyze the results of the bagging, random forest, and gradient boosting classification trees. Table 4.10 presents the overall variable importance in each model. The j variable can be said to be important when the total amount of Gini index is decreased by splits. After identifying the clusters and building the supervised learning, we can see from the Table 4.10 that Income has the highest variable importance for all models which is followed by household registration (HR) with 88.77, having formal loan with 65.19, Fin.Knowledge with 54.23, education with 48.31, and age with 22.46.

	Province	Region	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
1	Anhui	East	0%	16%	3%	0%	1%	80%
2	Beijing	East	1%	19%	28%	6%	18%	28%
3	Chongqing	East	0%	19%	3%	0%	0%	78%
4	Fujian	East	1%	29%	10%	2%	4%	55%
5	Gansu	East	0%	17%	2%	0%	1%	80%
6	Guangdong	East	1%	22%	9%	2%	5%	61%
7	Guangxi	East	0%	21%	4%	0%	1%	73%
8	Guizhou	East	1%	16%	3%	1%	1%	78%
9	Hainan	East	0%	14%	3%	1%	1%	82%
10	Hebei	East	0%	17%	3%	0%	1%	78%
11	Heilongjiang	East	0%	14%	2%	0%	0%	83%
12	Henan	Center	0%	18%	3%	0%	1%	78%
13	Hubei	Center	0%	20%	5%	0%	1%	73%
14	Hunan	Center	0%	19%	4%	1%	1%	76%
15	Inner	Center	0%	19%	5%	0%	2%	74%
16	Jiangsu	Center	0%	29%	13%	1%	4%	52%
17	Jiangxi	Center	0%	20%	5%	1%	1%	73%
18	Jilin	Center	0%	11%	2%	0%	0%	87%
19	Liaoning	Center	0%	18%	3%	0%	1%	78%
20	Ningxia	West	0%	21%	3%	0%	1%	75%
21	Qinghai	West	0%	15%	4%	0%	1%	80%
22	Shaanxi	West	0%	19%	3%	0%	1%	76%
23	Shandong	West	0%	26%	6%	0%	1%	67%
24	Shanghai	West	0%	35%	33%	3%	11%	18%
25	Shanxi	West	0%	14%	2%	0%	1%	82%
26	Sichuan	West	0%	20%	4%	0%	1%	74%
27	Tianjin	West	0%	39%	17%	1%	4%	40%
28	Yunnan	West	0%	17%	4%	0%	1%	77%
29	Zhejiang	West	1%	29%	14%	1%	4%	51%

Table 4.8: Cluster distribution by province

4.5.2.1 Comparing Prediction Area Under the Curve (AUC)

A receiver operating characteristic curve, known as ROC curve, is a one by one square graph. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at each threshold and visualize trade-offs between two extremes models (100%tue positive rate vs. 0% false positive rate). The default prediction accuracy of the models in terms of the standard area under the curve (AUC) derived from the receiver operator characteristic (ROC) curve. Where $0 \leq$ AUC ≤ 1 and AUC $= 0.5$ is random guessing, AUC $= 1$ model is perfectly fit, and AUC $= 0$ model is always wrong.

4.6 Micro Shinyapp

Micro Shinyapp[1] is a self explained application which provides an interactive user interface, using the 2015 CHFS data-set. Micro Shinyapp[1] is built to map the Cluster 1, Cluster 2, Cluster 3,

	Province	Region	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
1	Anhui	East	0%	2%	1%	0%	1%	3%
2	Beijing	East	9%	3%	11%	20%	21%	1%
3	Chongqing	East	2%	3%	1%	1%	0%	4%
4	Fujian	East	7%	6%	6%	9%	7%	3%
5	Gansu	East	0%	2%	1%	0%	1%	3%
6	Guangdong	East	31%	8%	10%	23%	15%	7%
7	Guangxi	East	3%	2%	1%	1%	1%	2%
8	Guizhou	East	4%	2%	1%	1%	1%	2%
9	Hainan	East	0%	2%	1%	2%	0%	3%
10	Hebei	East	1%	4%	2%	0%	1%	5%
11	Heilongjiang	East	2%	2%	1%	1%	0%	4%
12	Henan	Center	2%	2%	1%	1%	1%	3%
13	Hubei	Center	3%	4%	3%	2%	2%	4%
14	Hunan	Center	3%	4%	2%	2%	2%	4%
15	Inner	Center	0%	1%	1%	1%	1%	1%
16	Jiangsu	Center	2%	6%	8%	5%	7%	3%
17	Jiangxi	Center	0%	2%	1%	1%	0%	2%
18	Jilin	Center	0%	2%	1%	1%	1%	5%
19	Liaoning	Center	3%	5%	2%	1%	1%	6%
20	Ningxia	West	0%	1%	1%	0%	1%	2%
21	Qinghai	West	2%	1%	1%	1%	1%	2%
22	Shaanxi	West	1%	3%	1%	1%	1%	4%
23	Shandong	West	3%	6%	4%	1%	2%	5%
24	Shanghai	West	3%	5%	15%	10%	14%	1%
25	Shanxi	West	1%	3%	1%	1%	1%	5%
26	Sichuan	West	4%	4%	3%	1%	2%	5%
27	Tianjin	West	1%	5%	6%	2%	4%	1%
28	Yunnan	West	2%	2%	1%	1%	1%	3%
29	Zhejiang	West	15%	8%	11%	10%	10%	4%

Note: Cluster distribution by province over its cluster.

Table 4.9: Distribution of each cluster

	MLM	BAG	RF
Gender	15.69	5.74	0.23
Married	21.24	0	0
Age	22.46	33.60	27.12
Employed	0	1.42	1.39
Education	48.31	13.92	23.68
Party	5.83	5.29	2.51
Fin.Knowledge	54.23	0.41	3.24
Fin.Inter	11.47	3.58	1.01
HR	88.77	6.75	33.91
Region	East: 87.38, Central: 33.88	9.62	20.01
Income	99.569	100	100
Loan.Type	Formal: 65.198, Informal: 9.42, Both: 56.01	21.44	24.48

Table 4.10: Overall variable importance for each model

Cluster 4, Cluster 5, and Cluster 6. Where Cluster 1 and Cluster 4 are less populated wealth households, Cluster 2 and Cluster 6 are highly populated poor households, and Cluster 3 and

Cluster 5 have slightly moderate wealth

Hence, selecting all the Clusters at once will not provide an intuitive information about the financial strength of the Chinese households, but the distribution of the survey population. Selecting only Cluster 2 and Cluster 6, policy makers can see the distribution of the poor households' on the map and design suitable programs, aiming the most poor ares primarily. The guideline is presented in Appendix C.2.

Map of the households can be found based on their cluster level under the 📖 Map tab. Clicking on the provinces, name and the survey population can be seen based on the chosen Cluster. Red areas are the most populated areas, where the histogram of the selected cluster population felt between 80%-100% interval, and light blue areas are the less populated areas where the interval 0%-20%. ⊞ Table tab prints out the pivot table of the 2015 CHFS data-set.

Conclusion

This study focuses on the clustering of Chinese households based on their financial strength. We implement first unsupervised, (K-means clustering) using the asset owning variable (Networth), and then we implemented supervised learning (i.e., bagging, random forest, and boosting) for prediction. Based on unsupervised learning, one can conclude that access to credit is already toward better off. In another word, access to formal loan in Cluster 1 and Cluster 4 are the highest, where the financial asset owning values and income is higher. Our findings confirm the results by Fungáčová et al. (2014) and Chen and Jin (2017), and indicate that overall access to finance is low. Additionally, formal financial inclusion is particularly constrained which is distributed economically advantaged groups as a consequence disadvantaged households can not contribute to economic development.

Financial inclusion remains a major issue, since loan applicants usually excluded if they can not full fill the minimum requirements (e.g., income, employment, and financial assets) which cause inequity to arise. An efficient way, to reduce the income inequity, to boost consumer spending, and to ensure economic growth, is to swift government spending toward social welfare programs (Gan et al., 2014). As Gan et al. (2014) mention in his book social welfare spending in China is about 12% where as in the US is reaches up to 36.6%.

Access to financial services and funding problems may have a negative impact on the finan-

cial sustainability for individuals (e.g., households, self-employed), growth, and income inequity (Claessens, 2006; Demirguc-Kunt and Levine, 2008). For a constant increase in economic growth government's aim should be toward finance economically disadvantaged groups and promote financially disadvantage groups. The government should design suitable programs, aiming poor households (Cluster 2 and Cluster 6).

Chapter 5

Conclusion

5.1 Summary

This thesis is motivated by the increase attention on access to finance. More specifically, characteristics of households. The thesis contributes to the literature on as follows:

In Chapter 2, we review the literature on "determinants of finance", using both bibliometric and content analysis. We examine 210 published English published articles for bibliometric analysis, retrieved from Web of Science (WoS). We discussed: (1) influential aspects of the literature such as countries (Table 2.2), affiliations (Table 2.3), journals (Table 2.4), authors (Table 2.5), and articles (Table 2.6 and trend papers (Table 2.2), (2) two main research streams (Table 2.8), using co-citation (Figure 2.3), co-word (Figure 2.5), and content analysis, (3) the network between the authors (Figure 2.4), (4) 13 future research questions (Table 2.7) that we found using bibliometric and content analysis, and (5) Biblio in Section 3 user interface for a reproducible study. We recommend repeating the citation analysis, and considering other databases (such as; Scorpus, Google Scholars, Science Direct, etc.) for future studies.

In Chapter 3, we study the determinants of "access to finance" in China and we can conclude that access to finance in China is low and it is mostly favor to economically advantaged households. Access to formal loan has a positive association between income, asset variables, financial knowledge, and financial intermediaries.

For each model, we compare 15 models by four data splits and benchmark with 3 financial asset variables, and we did that for linear, bagging, random forest, and boosting for both access to loan and its type. So in total, we set 120 models. Additionally, dividing the data-set as CCP.0 & CCP.1 has a better predictive power to explain the characteristics of Chinese Households.

In Chapter 4, we propose to group disadvantaged households by using K-means clustering then, we implement bagging, random forest, and gradient boosting for prediction. Our findings, lines

with Fungáčová et al. (2014) and Chen and Jin (2017), indicate that overall access to finance is low and formal financial inclusion is particularly constrained which is distributed economically advantaged groups. For a constant increase in economic growth government's aim should design suitable programs, aiming poor households which we can localized using Micro in section 4.1. Additionally, we provide the code for this study in github[1].

Nonetheless, there are several limitations with CHFS data-set. Firstly, credit used was reported based on the household level, and unbalance formal loan accessibility restricts our study. Secondly, the majority of the households were male-headed which does not reveal the real gender biases to access credit. Lastly, ethnicity, and the number of financial intermediates which are closed to household were not reported.

---

[1]See at: https://github.com/seymakalay

Bibliography

Acedo FJ, Barroso C, Casanueva C, Galán JL. 2006. Co-authorship in management and organizational studies: An empirical and network analysis. Journal of Management Studies **43**: 957–983. 5

Alesina AF, Lotti F, Mistrulli PE. 2013. Do women pay more for credit? evidence from italy. Journal of the European Economic Association **11**: 45–66. 26, 27

Ali M, Peerlings J, Zhang X. 2014. Clustering as an organizational response to capital market inefficiency: evidence from microenterprises in ethiopia. Small Business Economics **43**: 697–709. 64

Allen F, Demirguc-Kunt A, Klapper L, Peria MSM. 2012. The foundations of financial inclusion: Understanding ownership and use of formal accounts. The World Bank. 28

Amarasinghe U, Samad M, Anputhas M. 2005. Spatial clustering of rural poverty and food insecurity in sri lanka. Food Policy **30**: 493–509. 64

Aportela F. 1999. Effects of financial access on savings by low-income people. 21

Aria M, Cuccurullo C. 2017. bibliometrix: An r-tool for comprehensive science mapping analysis. Journal of Informetrics **11**: 959–975. 4, 6

Ashraf N, Karlan D, Yin W. 2010. Female empowerment: Impact of a commitment savings product in the philippines. World development **38**: 333–344. 21

Asiedu E, Freeman JA, Nti-Addae A. 2012. Access to credit by small businesses: How relevant are race, ethnicity, and gender? American Economic Review **102**: 532–37. 28

Asiedu E, Kalonda-Kanyama I, Ndikumana L, Nti-Addae A. 2013. Access to credit by firms in sub-saharan africa: how relevant is gender? American Economic Review **103**: 293–97. 27

Aterido R, Beck T, Iacovone L. 2013. Access to finance in sub-saharan africa: is there a gender gap? World development **47**: 102–120. 34

Baker HK, Kumar S, Pattnaik D. 2020. Twenty-five years of review of financial economics: A bibliometric overview. Review of Financial Economics **38**: 3–23. 4

Bank W. 2011. An Action Plan to Improve Access to Finance for SMEs, Brussels, Belgium. an action plan to improve access to finance for smes, brussels, belgium.
URL http://documents.worldbank.org/curated/en/801151468152092070/Financial-inclusion-strategies-reference-framework 22

Bank W. 2019. World Bank world bank.
URL https://www.worldbank.org/en/publication/gfdr 22

Barber BM, Odean T. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. The quarterly journal of economics **116**: 261–292. 26

Bauer S, et al. 2016. Does credit access affect household income homogeneously across different groups of credit recipients? evidence from rural vietnam. Journal of rural studies **47**: 186–203. 25

Baumeister RF, Leary MR. 1997. Writing narrative literature reviews. Review of general psychology **1**: 311–320. 3

Bavelas JB. 1978. The social psychology of citations. Canadian Psychological Review/Psychologie Canadienne **19**: 158. 20

Baydas MM, Meyer RL, Aguilera-Alfred N. 1994. Credit rationing in small-scale enterprises: Special microenterprise programmes in Ecuador. The Journal of Development Studies **31**: 279–288. 37

Beck T, De La Torre A. 2006. The basic analytics of access to financial services. The World Bank. 27

Beck T, Demirgüç-Kunt A, Levine R. 2007. Finance, inequality and the poor. Journal of economic growth **12**: 27–49. 30

Beck T, Demirgüç-Kunt A, Maksimovic V. 2004. Bank competition and access to finance: International evidence. Journal of Money, Credit and Banking : 627–648. 22

Beck T, Demirgüç-Kunt A, Maksimovic V. 2005. Financial and legal constraints to growth: does firm size matter? The Journal of finance **60**: 137–177. 30

Beck T, Demirgüç-Kunt A, Maksimovic V. 2008. Financing patterns around the world: Are small firms different? Journal of financial economics **89**: 467–487. 31

Beck T, Levine R, Loayza N. 2000. ,"finance and the sources of growth," journal of financial economics 58, 261-300. 22, 30

Bellucci A, Borisov A, Zazzaro A. 2010. Does gender matter in bank–firm relationships? evidence from small business lending. Journal of Banking & Finance **34**: 2968–2984. 26

BIS. 2011. An Action Plan to Improve Access to Finance for SMEs, Brussels, Belgium. an action plan to improve access to finance for smes, brussels, belgium.
URL https://www.bis.org/bcbs/publ/d457.htm 31

Blanchard L, Zhao B, Yinger J. 2008. Do lenders discriminate against minority and woman entrepreneurs? Journal of Urban Economics **63**: 467–497. 27, 28

Blanchflower DG, Levine PB, Zimmerman DJ. 2003. Discrimination in the small-business credit market. Review of Economics and Statistics **85**: 930–943. 27, 38

Bourdieu P, Farage S. 1994. Rethinking the state: Genesis and structure of the bureaucratic field. Sociological theory **12**: 1–18. 3

Brandolini A, Magri S, Smeeding TM. 2010. Asset-based measurement of poverty. Journal of Policy Analysis and Management **29**: 267–284. 41, 67

Breiman L, Friedman J, Stone CJ, Olshen RA. 1984. Classification and regression trees. CRC press. 45, 70, 71

Broadus R. 1987. Toward a definition of "bibliometrics". Scientometrics **12**: 373–379. 3, 4

Bruns V, Holland DV, Shepherd DA, Wiklund J. 2008. The role of human capital in loan officers' decision policies. Entrepreneurship Theory and Practice **32**: 485–506. 34

Brush CG. 1992. Research on women business owners: Past trends, a new perspective and future directions. Entrepreneurship theory and practice **16**: 5–30. 26

Callon M, Courtial JP, Turner WA, Bauin S. 1983. From translations to problematic networks: An introduction to co-word analysis. Information (International Social Science Council) **22**: 191–235. 6

Campello M, Graham JR, Harvey CR. 2010. The real effects of financial constraints: Evidence from a financial crisis. Journal of financial Economics **97**: 470–487. 22

Campero A, Kaiser K. 2013. Access to credit: Awareness and use of formal and informal credit institutions. Technical report, Working Papers. 21

Caner A, Wolff EN. 2004. Asset poverty in the united states, 1984–99: Evidence from the panel study of income dynamics. Review of income and Wealth **50**: 493–518. 41, 67

Carter S, Mwaura S, Ram M, Trehan K, Jones T. 2015. Barriers to ethnic minority and women's enterprise: Existing evidence, policy tensions and unsettled questions. International Small Business Journal **33**: 49–69. 34

Carter, SL and Anderson, Susan and Shaw, Eleanor. 2001. Women's business ownership: a review of the academic, popular and internet literature . 38

Cavalluzzo K, Wolken J. 2005. Small business loan turndowns, personal wealth, and discrimination. The Journal of Business **78**: 2153–2178. 28

Cavalluzzo KS, Cavalluzzo LC. 1998. Market structure and discrimination: The case of small businesses. Journal of Money, Credit and Banking : 771–792. 27

Cavalluzzo KS, Cavalluzzo LC, Wolken JD. 2002. Competition, small business financing, and discrimination: Evidence from a new survey. The Journal of Business **75**: 641–679. 27, 28

Cesaroni F. 2010. Donne imprenditrici e banche. le ragioni di un rapporto difficile . 25

Chen Z, Jin M. 2017. Financial inclusion in China: use of credit. Journal of family and economic issues **38**: 528–540. 21, 53, 78, 81

Chodorow-Reich G. 2014. The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. The Quarterly Journal of Economics **129**: 1–59. 22

Claessens S. 2006. Access to financial services: a review of the issues and public policy objectives. The World Bank Research Observer **21**: 207–240. 21, 29, 30, 31, 35, 37, 79

Claessens S, Feijen E, Laeven L. 2008. Political connections and preferential access to finance: The role of campaign contributions. Journal of financial economics **88**: 554–580. 29, 34

Cole RA, Wolken JD. 1995. Financial services used by small businesses: Evidence from the 1993 National Survey of Small Business Finances. Fed. Res. Bull. **81**: 629. 38

Coleman S, Robb A. 2009. A comparison of new firm financing by gender: evidence from the kauffman firm survey data. Small Business Economics **33**: 397. 27

Collins D, Morduch J, Rutherford S, Ruthven O. 2009. Portfolios of the poor: how the world's poor live on $2 a day. Princeton University Press. 60

Cropanzano R. 2009. Writing nonempirical articles for journal of management: General thoughts and suggestions. 3

Cui Y, Nahm D, Tani M. 2013. Self-employment in China: Are rural migrant workers and urban residents alike? . 36, 62

Cull R, Li W, Sun B, Xu LC. 2015. Government connections and financial constraints: Evidence from a large representative sample of chinese firms. Journal of Corporate Finance **32**: 271–294. 29

Davis J, Mengersen K, Bennett S, Mazerolle L. 2014. Viewing systematic reviews and meta-analysis in social research through different lenses. SpringerPlus **3**: 1–9. 3

de Araújo Lima PF, Crema M, Verbano C. 2020. Risk management in smes: a systematic literature review and future directions. European Management Journal **38**: 78–94. 4

Demirguc-Kunt A, Klapper L. 2012. Measuring financial inclusion: The global findex database. The World Bank. 21, 30, 36, 61, 62

Demirguc-Kunt A, Klapper L. 2013. Measuring financial inclusion: Explaining variation in use of financial services across and within countries. Brookings Papers on Economic Activity **2013**: 279–340. 21, 23, 62

Demirguc-Kunt A, Klapper L, Peria MSM, Allen F, Martinez Peria MS. 2016. The foundations of financial inclusion: Understanding ownership and use of formal accounts. Journal of financial intermediation **27**: 1–30. ISSN 1042-9573. 21, 23

Demirguc-Kunt A, Levine R. 2008. Finance, financial sector policies, and long-run growth. The World Bank. 21, 29, 35, 37, 79

Demirguc-Kunt A, Maksimovic V. 1998. Law, finance, and firm growth. Journal of Finance **53**: 2107–2137. 22

Demko GJ. 2018. Population, politics, and geography: a global perspective. In Reordering The World. Routledge, 141–153. 36, 61

Deng S, Meng Y. 2013. Financial access and economic participation of youth with disabilities in China: an exploratory study. China Journal of Social Work **6**: 177–189. 38

Desai VS, Crook JN, Overstreet Jr GA. 1996. A comparison of neural networks and linear scoring models in the credit union environment. European journal of operational research **95**: 24–37. 63

Dickson BJ, Rublee MR. 2000. Membership has its privileges: the socioeconomic characteristics of Communist Party members in urban China. Comparative Political Studies **33**: 87–112. 38

Diodato VP, Gellatly P. 2013. Dictionary of bibliometrics. Routledge. 4

Dohmen TJ, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG. 2005. Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey . 26

Du Rietz A, Henrekson M. 2000. Testing the female underperformance hypothesis. Small Business Economics **14**: 1–10. 26

Dupas R, Robinson J. 2009. Savings constraints and microenterprise deveiopment: Evidence from a field experiment.„. NBER Working Paper Series . 21

Durand D. 1941. Risk elements in consumer installment financing. National Bureau of Economic Research, New York. 63

Dwyer RE. 2018. Credit, debt, and inequality. Annual Review of Sociology **44**: 237–261. 34

Ellis K, Lemma A. 2010. Investigating the impact of access to financial services on household investment . 24

Faccio M. 2006. Politically connected firms. American economic review **96**: 369–386. 28, 38, 54

Fairlie RW, Robb AM. 2009. Gender differences in business performance: evidence from the characteristics of business owners survey. Small Business Economics **33**: 375. 26

Fernandes D, Lynch Jr JG, Netemeyer RG. 2014. Financial literacy, financial education, and downstream financial behaviors. Management Science **60**: 1861–1883. 34

Finardi U, Buratti A. 2016. Scientific collaboration framework of brics countries: an analysis of international coauthorship. Scientometrics **109**: 433–446. 5

Firth M, Lin C, Liu P, Wong SM. 2009. Inside the black box: Bank credit allocation in china's private sector. Journal of Banking & Finance **33**: 1144–1155. 29

Fisher RA. 1936. The use of multiple measurements in taxonomic problems. Annals of eugenics **7**: 179–188. 63

Fisman R. 2001. Estimating the value of political connections. American economic review **91**: 1095–1102. 29

Frye T, Zhuravskaya E. 2000. Rackets, regulation, and the rule of law. Journal of Law, Economics, and Organization **16**: 478–502. 28

Fungáčová Z, Weill L. 2015. Understanding financial inclusion in china. China Economic Review **34**: 196–206. 29, 62

Fungáčová Z, Weill L, et al. 2014. A view on financial inclusion in asian countries . 21, 23, 36, 37, 53, 54, 61, 62, 78, 81

Gale F, Callender R. 2006. New directions in China's agricultural lending. DIANE Publishing. 65

Gale HF, Lohmar B, Tuan FC. 2005. China's new farm subsidies. USDA-ERS WRS-05-01 . 64

Galiani S, Schargrodsky E. 2010. Property rights for the poor: Effects of land titling. Journal of Public Economics **94**: 700–729. ISSN 0047-2727. 25, 34

Gan L, Yin Z, Jia N, Xu S, Ma S, Zheng L. 2014. Data you need to know about China. Springer. 21, 30, 54, 78

Garfield E, Merton RK. 1979. Citation indexing: Its theory and application in science, technology, and humanities, volume 8. Wiley New York. 19

Geng N, N'Diaye P. 2012. Determinants of corporate investment in china: Evidence from cross-country firm level data. 12/80. International Monetary Fund. 29

Gerlach-Kristen P, O'Connell B, O'Toole C. 2015. Do credit constraints affect sme investment and employment? The Economic and Social Review **46**: 51–86. 22

Ghosh S, Vinod D. 2017. What constrains financial inclusion for women? evidence from indian micro data. World Development **92**: 60–81. ISSN 0305-750X. 27, 34

Giné X. 2011. Access to capital in rural Thailand: An estimated model of formal vs. informal credit. Journal of Development Economics **96**: 16–29. 38

Glänzel W. 2001. National characteristics in international scientific co-authorship relations. Scientometrics **51**: 69–115. 6

Glass GV. 1976. Primary, secondary, and meta-analysis of research. Educational researcher **5**: 3–8. 3

Grant C. 2007. Estimating credit constraints among us households. Oxford Economic Papers **59**: 583–605. 24

Gucz HN. 2014. How hidden fractures still threaten the world economy. 30

Hale G, Long C. 2011. Chapter 13" what are the sources of financing for chinese firms?". The Evolving Role of Asia in Global Finance (Frontiers of Economics and Globalization, 9) Emerald Group Publishing Limited : 313–339. 29, 63

Hannig A, Jansen S. 2010. Financial inclusion and financial stability: Current policy issues . 21, 23, 35, 60

Hartigan JA, Wong MA. 1979. Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics) **28**: 100–108. 69

Haveman R, Wolff EN. 2004. The concept and measurement of asset poverty: Levels, trends and composition for the us, 1983–2001. The Journal of Economic Inequality **2**: 145–169. 41, 67

Hay JR, Shleifer A. 1998. Private enforcement of public laws: A theory of legal reform. The American Economic Review **88**: 398–403. 28

Hayashi F. 1985. Tests for liquidity constraints: a critical survey. 22

Henninger N, Snel M. 2002. Where are the poor. Experiences with the development and use of poverty maps. Washington, DC: World Resources Institute . 63

Honohan P. 2004. Financial sector policy and the poor: Selected findings and issues. The World Bank. 30

Honohan P. 2008. Cross-country variation in household access to financial services. Journal of Banking & Finance **32**: 2493–2500. 30

Huang CL, Chen MC, Wang CJ. 2007. Credit scoring with a data mining approach based on support vector machines. Expert systems with applications **33**: 847–856. 63

Huang J, Jin M, Deng S, Guo B, Zou L, Sherraden M. 2013. Asset poverty in urban china: A study using the 2002 chinese household income project. Journal of Social Policy **42**: 763–781. 38, 41, 67

Hung C, Chen JH. 2009. A selective ensemble based on expected probabilities for bankruptcy prediction. Expert systems with applications **36**: 5297–5303. 63

Jaffee DM, Russell T. 1976. Imperfect information, uncertainty, and credit rationing. The Quarterly Journal of Economics **90**: 651–666. 22

Jain AK. 2010. Data clustering: 50 years beyond k-means. Pattern recognition letters **31**: 651–666. 69

James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning, volume 112. Springer. 42, 69

Jianakoplos NA, Bernasek A. 1998. Are women more risk averse? Economic inquiry **36**: 620–630. 26

Johnson S, Mitton T. 2003. Cronyism and capital controls: evidence from Malaysia. Journal of financial economics **67**: 351–382. 29, 38

Källestål C, Blandón EZ, Peña R, Peréz W, Contreras M, Persson LÅ, Sysoev O, Selling KE. 2020. Assessing the multiple dimensions of poverty. data mining approaches to the 2004–14 health and demographic surveillance system in cuatro santos, nicaragua. Frontiers in Public Health **7**: 409. 64

Kanbur R, Zhang XB. 1998. Which regional inequality? the evolution of rural-urban and inland-coastal inequality in china, 1983-1995 . 25

Kessler MM. 1963. Bibliographic coupling between scientific papers. American documentation **14**: 10–25. 4, 14

Khwaja AI, Mian A. 2005. Do lenders favor politically connected firms? rent provision in an emerging financial market. The Quarterly Journal of Economics **120**: 1371–1411. 38, 54

Kim J, McMillan SJ. 2008. Evaluation of internet advertising research: A bibliometric analysis of citations from key sources. Journal of Advertising **37**: 99–112. 6

Kira AR. 2013. The evaluation of the factors influence the access to debt financing by tanzanian smes . 24

Klapper LF, Parker SC. 2011. Gender and the business environment for new firm creation. The World Bank Research Observer **26**: 237–257. 25

Kochen M. 1974. Principles of information retrieval. Melville Pub. Co. 20

Kuhn M. 2008. Building predictive models in r using the caret package. Journal of statistical software **28**: 1–26. 69

Kumhof M, Rancière R, Winant P. 2015. Inequality, leverage, and crises. American Economic Review **105**: 1217–45. 31

Kunisch S, Menz M, Bartunek JM, Cardinal LB, Denyer D. 2018. Feature topic at organizational research methods: How to conduct rigorous and impactful literature reviews? Organizational Research Methods **21**: 519–523. 3

Laloë F, Mosseri R. 2009. Bibliometric evaluation of individual researchers: not even right... not even wrong! Europhysics News **40**: 26–29. 19

Levine R. 2004. Finance and growth: Theory, evidence, and mechanism. Hanbook of Economic Review **88**: 3. 30

Li H, Liu PW, Zhang J, Ma N. 2007. Economic returns to communist party membership: Evidence from urban Chinese twins. The Economic Journal **117**: 1504–1520. 38

Li H, Meng L, Wang Q, Zhou LA. 2008. Political connections, financing and firm performance: Evidence from Chinese private firms. Journal of development economics **87**: 283–299. 21, 29, 38, 61

Li K, Zhao C. 2011. Determinants of Self-employment in China: Evidence from Cross-regional Data. China & World Economy **19**: 49–67. 65

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology **62**: e1–e34. 3

Liu W, Gu M, Hu G, Li C, Liao H, Tang L, Shapira P. 2014. Profile of developments in biomass-based bioenergy research: a 20-year perspective. Scientometrics **99**: 507–521. 3

Liu X, Bollen J, Nelson ML, Van de Sompel H. 2005. Co-authorship networks in the digital library research community. Information processing & management **41**: 1462–1480. 6

Loscocco KA, Robinson J, Hall RH, Allen JK. 1991. Gender and small business success: An inquiry into women's relative disadvantage. Social forces **70**: 65–85. 26

Maksimovic V, Demirgüç-Kunt A. 1996. Institutions, financial markets and firm debt maturity. World Bank Policy Research Working Paper . 28

McMillan J, Woodruff C. 1999. Interfirm relationships and informal credit in vietnam. The Quarterly Journal of Economics **114**: 1285–1320. 28

McParland D, Gormley IC, McCormick TH, Clark SJ, Kabudula CW, Collinson MA. 2014. Clustering south african households based on their asset status using latent variable models. The annals of applied statistics **8**: 747. 64

Merediz-Sola I, Bariviera AF. 2019. A bibliometric analysis of bitcoin scientific production. Research in International Business and Finance . 5

Moher D, Liberati A, Tetzlaff J, Altman DG, et al. 2010. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Int J Surg **8**: 336–341. 3

Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. 2000. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. Bmj **320**: 686–690. 63

Morduch J, Haley B. 2002. Analysis of the effects of microfinance on poverty reduction. New York: NYU Wagner Working Paper **1014**. 30

Morsy H. 2020. Access to finance–mind the gender gap. The Quarterly Review of Economics and Finance **78**: 12–21. 27

Munnell AH, Tootell GM, Browne LE, McEneaney J. 1996. Mortgage lending in boston: Interpreting hmda data. The American Economic Review : 25–53. 28

Muravyev A, Talavera O, Schäfer D. 2009. Entrepreneurs' gender and financial constraints: Evidence from international data. Journal of Comparative Economics **37**: 270–286. 26

Mushinski DW. 1999. An analysis of offer functions of banks and credit unions in Guatemala. The journal of Development studies **36**: 88–112. 38

Mysiak J, Giupponi C, Rosato P. 2005. Towards the development of a decision support system for water resource management. Environmental Modelling & Software **20**: 203–214. 63

Nguyen HT, Nguyen HM, Troege M, Nguyen AT. 2020. Debt aversion, education, and credit self-rationing in smes. Small Business Economics : 1–19. 24

ONeil C. 2017. Weapons of math destruction: how big data increases inequality and threatens democracy. 63

Orens R, Reheul AM. 2013. Do ceo demographics explain cash holdings in smes? European Management Journal **31**: 549–563. 24

Peters H, Van Raan A. 1991. Structuring scientific activities by co-author analysis: An expercise on a university faculty level. Scientometrics **20**: 235–255. 6

Pinto CF, Serra FR, Ferreira MP. 2014. A bibliometric study on culture research in international business. BAR-Brazilian Administration Review **11**: 340–363. 3

Popov A, Rocholl J. 2018. Do credit shocks affect labor demand? evidence for employment and wages during the financial crisis. Journal of Financial Intermediation **36**: 16–27. 22

Porta RL, Lopez-de Silanes F, Shleifer A, Vishny RW. 1998. Law and finance. Journal of political economy **106**: 1113–1155. 28

Pritchard A, et al. 1969. Statistical bibliography or bibliometrics. Journal of documentation **25**: 348–349. 3, 4

Ramalho R. 2007. The persistence of corruption: evidence from the 1992 presidential impeachment in brazil. Technical report, The World Bank. 29, 38

Research A, Consulting L. 2009. Financial capability in the united states: Initial report of research findings from the 2009 national survey. 30

Reuters T. 2008. Whitepaper using bibliometrics: A guide to evaluating research performance with citation data, 2008. acesso em: 18/03/2014. 19

Riley Jr WB, Chow KV. 1992. Asset allocation and individual risk aversion. Financial Analysts Journal **48**: 32–37. 24

Robb A, Wolken JD. 2002. Firm, owner, and financing characteristics: Differences between female- and male-owned small businesses . 26

Romer PM. 1990. Endogenous technological change. Journal of political Economy **98**: S71–S102. 24

Rosa P, Carter S, Hamilton D. 1996. Gender as a determinant of small business performance: Insights from a british study. Small business economics **8**: 463–478. 26

Rui L, Xi Z. 2010. Econometric analysis of credit constraints of Chinese rural households and welfare loss. Applied Economics **42**: 1615–1625. 39

Schebesch KB, Stecking R. 2005. Support vector machines for classifying and describing credit applicants: detecting typical and critical regions. Journal of the Operational Research Society **56**: 1082–1088. 63

Schneider JW, Aagaard K, Bloch CW. 2016. What happens when national research funding is linked to differentiated publication counts? a comparison of the australian and norwegian publication-based funding models. Research Evaluation **25**: 244–256. 19

Schumpeter JA. 2013. Capitalism, socialism and democracy. routledge. 30

Shi Y, Li X. 2019. A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms. Heliyon **5**: e02997. 4

Shoji M, Aoyagi K, Kasahara R, Sawada Y, Ueyama M. 2012. Social capital formation and credit access: Evidence from Sri Lanka. World Development **40**: 2522–2536. 29, 38, 54

Small H. 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for information Science **24**: 265–269. 4

Snyder H. 2019. Literature review as a research methodology: An overview and guidelines. Journal of Business Research **104**: 333–339. 3

Solow RM. 1956. A contribution to the theory of economic growth. The quarterly journal of economics **70**: 65–94. 24

Sparreboom P, Duflos E. 2012. Financial inclusion in the people's republic of china: An analysis of existing research and public data. China Papers on Inclusiveness **7**: 11–64. 35, 60, 61

Stefani ML, Vacca V. 2015. Small firms' credit access in the euro area: Does gender matter? CESifo Economic Studies **61**: 165–201. 26

Stiglitz JE, Weiss A. 1981. Credit rationing in markets with imperfect information. The American Economic Review **71**: 393–410. ISSN 00028282.
URL http://www.jstor.org/stable/1802787 22

Tang S, Guo S. 2017. Formal and informal credit markets and rural credit demand in China. In 2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS). IEEE, 1–7. 39

Teixeira AA, Carvalho L. 2014. Where are the poor in mainstream international economics? Poverty & Public Policy **6**: 215–238. 4

Tejerina L, Westley GD, Tejerina L, Westley GD, Tejerina L, Westley GD, Tejerina L, Westley GD. 2007. Financial services for the poor: household survey sources and gaps in borrowing and saving . 21

Turvey CG, Kong R. 2010. Informal lending amongst friends and relatives: can microcredit compete in rural China? China Economic Review **21**: 544–556. 38

Ullah Bhuiyan H, Shah A. 2011. A literature survey of value creation through relationship banking. African Journal of Business Management **5**: 13517–13522. 34

UNCTAD S. 2005. Improving the competitiveness of smes through enhancing productive capacity. Proceedings of Four Expert Meetings. 64

Wan G. 2008. Inequality and growth in modern China. OUP Oxford. 35, 60

Wan G, Sebastian I. 2011. Poverty in asia and the pacific: An update. Asian Development Bank Economics Working Paper . 60

Wang C, Wan G, Yang D. 2014. Income inequality in the people's republic of china: Trends, determinants, and proposed remedies. Journal of Economic Surveys **28**: 686–708. 61

Watson J. 2002. Comparing the performance of male-and female-controlled businesses: relating outputs to inputs. Entrepreneurship theory and practice **26**: 91–100. 26

West D. 2000. Neural network credit scoring models. Computers & Operations Research **27**: 1131–1152. 63

Wolman H, Hincapie D. 2010. Clusters and cluster-based development: A literature review and policy discussion. George Washington Institute of Public Policy (GWIPP), Washington, DC . 63

Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. 2013. Rameses publication standards: Meta-narrative reviews. Journal of Advanced Nursing **69**: 987–1004. 3

Xu L, Clarke G, Zou Hf. 2003. Finance and income inequality: Test of alternative theories. The World Bank. 22, 31

Yoshimura M, Fujimi Y, Izui K, Nishiwaki S. 2006. Decision-making support system for human resource allocation in product development projects. International journal of production research **44**: 831–848. 63

Young A. 2013. Inequality, the urban-rural gap, and migration. The Quarterly Journal of Economics **128**: 1727–1785. 25

Yuan Y, Xu L. 2015. Are poor able to access the informal credit market? Evidence from rural households in China. China Economic Review **33**: 232–246. 37, 38

Yunus M. 1998. Banker to the Poor. Penguin Books India. 65

Zeller M. 1994. Determinants of credit rationing: A study of informal lenders and formal credit groups in Madagascar. World development **22**: 1895–1907. 37

Zhai Q, Su J, Ye M. 2014. Focus on china: the current status of entrepreneurship research in china. Scientometrics **98**: 1985–2006. 5

Zhuang J. 2008. Inclusive growth toward a harmonious society in the people's republic of china: policy implications. Asian Development Review **25**: 22. 61

Zupic I, Čater T. 2015. Bibliometric methods in management and organization. Organizational Research Methods **18**: 429–472. 4, 6

Appendix A

Bibliometric Dataset of Chapter 2

A.1   Bibliometric Data-set

Table A.1: Papers are used for bibliometric analysis

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
1	Abor J, 2006, Environ Plann C	How Are Smes Financed? Evidence From The Ghanaian Nontraditional Export Sector	0	0.00	5	0.33
2	Adetunji Om, 2019, J Int Dev	The Relative Impact Of Income And Financial Literacy On Financial Inclusion In Nigeria	0	0.00	2	1.00
3	Agier I, 2013, Small Bus Econ	Subjectivity In Credit Allocation To Micro-Entrepreneurs: Evidence From Brazil	2	0.25	23	2.88
4	Agier I, 2013, World Dev	Microfinance And Gender: Is There A Glass Ceiling On Loan Size?	5	0.62	73	9.12
5	Al-Bahrani A, 2019, J Consum Aff	Racial Differences In The Returns To Financial Literacy Education	0	0.00	12	6.00
6	Alden L, 2016, Kyklos	Discrimination In The Credit Market? Access To Financial Capital Among Self-Employed Immigrants	0	0.00	23	4.60
7	Alesina Af, 2013, J Eur Econ Assoc	Do Women Pay More For Credit? Evidence From Italy	11	1.38	81	10.12
8	Alhassan A, 2021, Int J Financ Econ	The Relationship Between Political Instability And Financial Inclusion: Evidence From Middle East And North Africa	0		2	Inf
9	Allen F, 2014, J Afr Econ	The African Financial Development And Financial Inclusion Gaps	1	0.14	57	8.14
10	Allen F, 2016, J Financ Intermed	The Foundations Of Financial Inclusion: Understanding Ownership And Use Of Formal Accounts	13	2.60	118	23.60
11	Anarfo Eb, 2020, Res Int Bus Financ	Financial Regulation And Financial Inclusion In Sub-Saharan Africa: Does Financial Stability Play A Moderating Role?	0	0.00	10	10.00
12	Anderson Cl, 2002, World Dev	Microcredit, Social Capital, And Common Pool Resources	0	0.00	101	5.32
13	Aristei D, 2016, Financ Res Lett	Does Gender Matter For Firms' Access To Credit? Evidence From International Data	1	0.20	17	3.40
14	Asiedu E, 2012, Am Econ Rev	Access To Credit By Small Businesses: How Relevant Are Race, Ethnicity, And Gender?	7	0.78	37	4.11
15	Asiedu E, 2013, Am Econ Rev	Access To Credit By Firms In Sub-Saharan Africa: How Relevant Is Gender?	9	1.12	44	5.50
16	Aterido R, 2013, World Dev	Access To Finance In Sub-Saharan Africa: Is There A Gender Gap?	16	2.00	83	10.38
17	Bailey W, 2011, J Financ Quant Anal	Bank Loans With Chinese Characteristics: Some Evidence On Inside Debt In A State-Controlled Banking System	0	0.00	79	7.90
18	Banerjee Av, 2014, Rev Econ Stud	Do Firms Want To Borrow More? Testing Credit Constraints Using A Directed Lending Program	1	0.14	104	14.86

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
19	Bates T, 1997, J Urban Aff	Unequal Access: Financial Institution Lending To Black- And White-Owned Small Business Start-Ups	1	0.04	24	1.00
20	Bates T, 2016, Econ Dev Q	Impacts Of Owner Race And Geographic Context On Access To Small-Business Financing	0	0.00	14	2.80
21	Baydas Mm, 1994, World Dev	Discrimination Against Women In Formal Credit Markets - Reality Or Rhetoric	3	0.11	19	0.70
22	Bayer P, 2017, J Urban Econ	Racial And Ethnic Price Differentials In The Housing Market	0	0.00	9	2.25
23	Bayer P, 2018, Rev Financ Stud	What Drives Racial And Ethnic Differences In High-Cost Mortgages? The Role Of High-Risk Lenders	0	0.00	16	5.33
24	Beck T, 2006, J Bank Financ	Small And Medium-Size Enterprises: Access To Finance As A Growth Constraint	8	0.53	626	41.73
25	Beck T, 2008, J Financ Econ	Financing Patterns Around The World: Are Small Firms Different?	4	0.31	360	27.69
26	Beck T, 2008, World Bank Econ Rev	Access To Finance: An Unfinished Agenda	8	0.62	96	7.38
27	Beck T, 2014, J Afr Econ	Sme Finance In Africa	0	0.00	17	2.43
28	Beck T, 2018, J Bank Financ	Sex And Credit: Do Gender Interactions Matter For Credit Market Outcomes?	1	0.33	12	4.00
29	Berkovec Ja, 1994, J Real Estate Financ	Race, Redlining, And Residential Mortgage Loan Performance	8	0.30	70	2.59
30	Biggs T, 2002, J Econ Behav Organ	Ethnic Networks And Access To Credit: Evidence From The Manufacturing Sector In Kenya	1	0.05	41	2.16
31	Bigsten A, 2006, World Bank Res Obser	What Have We Learned From A Decade Of Manufacturing Enterprise Surveys In Africa?	1	0.07	67	4.47
32	Black Ha, 2001, Rev Black Polit Econ	Do Lenders Discriminate Against Low-Income Borrowers?	0	0.00	2	0.10
33	Black Ha, 2003, J Bank Financ	Is There Discrimination In Mortgage Pricing? The Case Of Overages	6	0.33	24	1.33
34	Blanchard L, 2008, J Urban Econ	Do Lenders Discriminate Against Minority And Woman Entrepreneurs?	11	0.85	74	5.69
35	Blanchflower Dg, 2003, Rev Econ Stat	Discrimination In The Small-Business Credit Market	31	1.72	237	13.17
36	Boehm Tp, 2006, Hous Policy Debate	Rates And Race: An Analysis Of Racial Disparities In Mortgage Rates	1	0.07	24	1.60
37	Bongomin Goc, 2018, Int J Emerg Mark	Analyzing The Relationship Between Institutional Framework And Financial Inclusion In Rural Uganda: A Social Network Perspective	0	0.00	5	1.67
38	Brown M, 2011, Econ Policy	Who Needs Credit And Who Gets Credit In Eastern Europe?	2	0.20	78	7.80
39	Browne Le, 1995, New Engl Econ Rev	Mortgage Lending In Boston - A Response To The Critics	6	0.23	35	1.35
40	Bruder J, 2011, Int J Entrep Behav R	Financial Constraints Of Ethnic Entrepreneurship: Evidence From Germany	0	0.00	24	2.40
41	Calern Ps, 2004, Hous Policy Debate	Neighborhood Patterns Of Subprime Lending: Evidence From Disparate Cities	1	0.06	53	3.12
42	Campbell Jy, 2006, J Financ	Household Finance	1	0.07	769	51.27
43	Carrillo P, 2009, Reg Sci Urban Econ	Alternative Measures Of Homeownership Gaps Across Segregated Neighborhoods	0	0.00	11	0.92

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
44	Carter S, 2007, Entrep Theory Pract	Gender, Entrepreneurship, And Bank Lending: The Criteria And Processes Used By Bank Loan Officers In Assessing Applications	7	0.50	188	13.43
45	Carter S, 2015, Int Small Bus J	Barriers To Ethnic Minority And Women's Enterprise: Existing Evidence, Policy Tensions And Unsettled Questions	2	0.33	95	15.83
46	Cavalluzzo K, 2005, J Bus	Small Business Loan Turndowns, Personal Wealth, And Discrimination	9	0.56	111	6.94
47	Cavalluzzo Ks, 1998, J Money Credit Bank	Market Structure And Discrimination: The Case Of Small Businesses	26	1.13	111	4.83
48	Cavalluzzo Ks, 2002, J Bus	Competition, Small Business Financing, And Discrimination: Evidence From A New Survey	27	1.42	186	9.79
49	Chakravarty Sr, 2013, J Policy Model	Financial Inclusion In India: An Axiomatic Approach	3	0.38	60	7.50
50	Chandio Aa, 2018, Emerg Mark Financ Tr	Determinants Of Credit Constraints: Evidence From Sindh, Pakistan	2	0.67	7	2.33
51	Charles Kk, 2002, Rev Econ Stat	The Transition To Home Ownership And The Black-White Wealth Gap	3	0.16	99	5.21
52	Charumilind C, 2006, J Bus	Connected Lending: Thailand Before The Financial Crisis	8	0.53	182	12.13
53	Chaudhuri K, 2020, Small Bus Econ	Gender, Small Firm Ownership, And Credit Access: Some Insights From India	1	1.00	9	9.00
54	Chen X, 2020, J Bank Financ	Gender Gap In Peer-To-Peer Lending: Evidence From China	0	0.00	9	9.00
55	Chen Ys, 2014, J Financ Serv Res	The Benefits Of Political Connection: Evidence From Individual Bank-Loan Contracts	1	0.14	35	5.00
56	Chen Zb, 2017, J Fam Econ Iss	Financial Inclusion In China: Use Of Credit	1	0.25	9	2.25
57	Chibba M, 2009, Eur J Dev Res	Financial Inclusion, Poverty Reduction And The Millennium Development Goals	0	0.00	87	7.25
58	Claessens S, 2007, J Comp Econ	Finance And Inequality: Channels And Evidence	4	0.29	154	11.00
59	Claessens S, 2008, J Financ Econ	Political Connections And Preferential Access To Finance: The Role Of Campaign Contributions	12	0.92	557	42.85
60	Cloud C, 1993, Rev Black Polit Econ	What Do We Know About Racial-Discrimination In Mortgage Markets	0	0.00	17	0.61
61	Cole R, 2016, J Financ Stabil	Who Needs Credit And Who Gets Credit? Evidence From The Surveys Of Small Business Finances	2	0.40	36	7.20
62	Coleman S, 2000, J Small Bus Manage	Access To Capital And Terms Of Credit: A Comparison Of Men- And Women-Owned Small Businesses	15	0.71	232	11.05
63	Corrado G, 2015, J Econ Geogr	The Geography Of Financial Inclusion Across Europe During The Global Crisis	1	0.17	20	3.33
64	Corsi M, 2017, J Dev Stud	Gender Discrimination In Microfinance? Some Evidence From Uganda	1	0.25	3	0.75
65	Courchane Mj, 2007, J Real Estate Res	The Pricing Of Home Mortgage Loans To Minority Borrowers: How Much Of The Apr Differential Can We Explain?	2	0.14	34	2.43
66	Cozarenco A, 2018, J Bus Ethics	Gender Biases In Bank Lending: Lessons From Microcredit In France	1	0.33	12	4.00
67	Cull R, 2015, J Corp Financ	Government Connections And Financial Constraints: Evidence From A Large Representative Sample Of Chinese Firms	2	0.33	132	22.00

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
68	Cull R, 2019, Oxford B Econ Stat	Dual Credit Markets And Household Usage To Finance: Evidence From A Representative Chinese Household Survey	0	0.00	1	0.50
69	Cuong Nv, 2008, Dev Econ	Is A Governmental Micro-Credit Program For The Poor Really Pro-Poor? Evidence From Vietnam	1	0.08	41	3.15
70	Davutyan N, 2016, Emerg Mark Financ Tr	Determinants Of Saving-Borrowing Decisions And Financial Inclusion In A High Middle Income Country: The Turkish Case	2	0.40	8	1.60
71	Deku Sy, 2016, Eur J Financ	Access To Consumer Credit In The Uk	2	0.40	8	1.60
72	Demirguec-Kunt A, 2013, Brookings Pap Eco Ac	Measuring Financial Inclusion: Explaining Variation In Use Of Financial Services Across And Within Countries	9	1.12	102	12.75
73	Duy Vq, 2012, Afr Asian Stud	Determinants Of Household Access To Formal Credit In The Rural Areas Of The Mekong Delta, Vietnam	1	0.11	12	1.33
74	Dymski G, 2013, Fem Econ	Race, Gender, Power, And The Us Subprime Mortgage And Foreclosure Crisis: A Meso Analysis	1	0.12	39	4.88
75	Eddleston Ka, 2016, Entrep Theory Pract	Do You See What I See? Signaling Effects Of Gender And Firm Characteristics On Financing Entrepreneurial Ventures	1	0.20	62	12.40
76	Ezeala-Harrison F, 2008, J Econ Issues	Determinants Of Housing Loan Patterns Toward Minority Borrowers In Mississippi	0	0.00	2	0.15
77	Fafchamps M, 2000, J Dev Econ	Ethnicity And Credit In African Manufacturing	3	0.14	122	5.81
78	Fang L, 2020, Real Estate Econ	Gender Equality In Mortgage Lending	0	0.00	1	1.00
79	Fay M, 1993, J Bus Venturing	Gender Bias And The Availability Of Business Loans	8	0.29	105	3.75
80	Fernandes D, 2014, Manage Sci	Financial Literacy, Financial Education, And Downstream Financial Behaviors	2	0.29	439	62.71
81	Firestone S, 2014, J Money Credit Bank	Race, Ethnicity, And Credit Card Marketing	0	0.00	7	1.00
82	Firth M, 2009, J Bank Financ	Inside The Black Box: Bank Credit Allocation In China's Private Sector	3	0.25	200	16.67
83	Fletschner D, 2009, World Dev	Rural Women's Access To Credit: Market Imperfections And Intrahousehold Dynamics	5	0.42	48	4.00
84	Fletschner D, 2010, J Dev Stud	Risk, Credit Constraints And Financial Efficiency In Peruvian Agriculture	2	0.18	20	1.82
85	Fraser S, 2009, Int Small Bus J	Is There Ethnic Discrimination In The Uk Market For Small Business Credit?	2	0.17	41	3.42
86	Freeland Re, 2016, J Small Bus Manage	How Does Race And Ethnicity Affect Persistence In Immature Ventures?	0	0.00	17	3.40
87	Fu Jt, 2017, J Asian Econ	Can Firms With Political Connections Borrow More Than Those Without? Evidence From Firm-Level Data For Indonesia	0	0.00	7	1.75
88	Fungacova Z, 2015, China Econ Rev	Understanding Financial Inclusion In China	9	1.50	64	10.67
89	Gabriel Sa, 2005, J Urban Econ	Homeownership In The 1980s And 1990s: Aggregate Trends And Racial Gaps	1	0.06	57	3.56
90	Galli E, 2020, Eur Financ Manag	Bank Credit Constraints For Women-Led Smes: Self-Restraint Or Lender Bias?	0	0.00	1	1.00
91	Garikipati S, 2017, J Dev Stud	Microfinance And Gender: Issues, Challenges And The Road Ahead	0	0.00	17	4.25

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
92	Ghent Ac, 2014, Reg Sci Urban Econ	Differences In Subprime Loan Pricing Across Races And Neighborhoods	3	0.43	18	2.57
93	Ghosh S, 2017, World Dev	What Constrains Financial Inclusion For Women? Evidence From Indian Micro Data	5	1.25	31	7.75
94	Grant C, 2007, Oxford Econ Pap	Estimating Credit Constraints Among Us Households	1	0.07	21	1.50
95	Grohmann A, 2018, World Dev	Does Financial Literacy Improve Financial Inclusion? Cross Country Evidence	2	0.67	50	16.67
96	Hagstrom P, 2021, J Int Dev	Financial Inclusion Of Individuals Who Arrived As Refugees To The United States	0		0	
97	Han S, 2004, J Real Estate Financ	Discrimination In Lending: Theory And Evidence	3	0.18	13	0.76
98	Hansen H, 2014, Econ Lett	Estimates Of Gender Differences In Firm's Access To Credit In Sub-Saharan Africa	2	0.29	16	2.29
99	Hansen H, 2014, J Dev Stud	The Myth Of Female Credit Discrimination In African Manufacturing	5	0.71	25	3.57
100	Hanson A, 2016, J Urban Econ	Discrimination In Mortgage Lending: Evidence From A Correspondence Experiment	2	0.40	34	6.80
101	Harhoff D, 1998, J Bank Financ	Lending Relationships In Germany - Empirical Evidence From Survey Data	2	0.09	270	11.74
102	Harkness Sk, 2016, Soc Psychol Quart	Discrimination In Lending Markets: Status And The Intersections Of Gender And Race	0	0.00	19	3.80
103	Hastings Js, 2013, Annu Rev Econ	Financial Literacy, Financial Education, And Economic Outcomes	2	0.25	198	24.75
104	Heikkila A, 2016, J Dev Stud	Social Capital And Access To Credit: Evidence From Uganda	0	0.00	11	2.20
105	Hernandez-Canovas G, 2010, Small Bus Econ	Relationship Lending And Sme Financing In The Continental European Bank-Based System	1	0.09	74	6.73
106	Holmes A, 1994, J Financ	Mortgage Redlining - Race, Risk, And Demand	4	0.15	43	1.59
107	Horska E, 2013, Agr Econ-Czech	Knowledge And Financial Skills As The Factors Determining The Financial Exclusion Process Of Rural Dwellers In Poland	0	0.00	5	0.62
108	Houston Jf, 2014, J Account Res	Political Connections And The Cost Of Bank Loans	1	0.14	172	24.57
109	Hung Chd, 2017, J Financ Stabil	Bank Political Connections And Performance In China	0	0.00	21	5.25
110	Hussain J, 2019, J Int Dev	Gender, Microcredit And Poverty Alleviation In A Developing Country: The Case Of Women Entrepreneurs In Pakistan	0	0.00	7	3.50
111	Infante L, 2014, J Corp Financ	Political Connections And Preferential Lending At Local Level: Some Evidence From The Italian Credit Market	1	0.14	36	5.14
112	Johnson S, 2011, J Dev Stud	Financial Access And Exclusion In Kenya And Uganda	5	0.50	37	3.70
113	Kairiza T, 2017, Small Bus Econ	Gender Differences In Financial Inclusion Amongst Entrepreneurs In Zimbabwe	1	0.25	18	4.50
114	Kara A, 2017, J Financ Serv Res	Household Access To Mortgages In The Uk	0	0.00	2	0.50
115	Karakurum-Ozdemir K, 2019, Soc Indic Res	Financial Literacy In Developing Countries	0	0.00	12	6.00
116	Kau Jb, 2012, J Real Estate Financ	Racial Discrimination And Mortgage Lending	1	0.11	14	1.56

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
117	Khwaja Ai, 2005, Q J Econ	Do Lenders Favor Politically Connected Firms? Rent Provision In An Emerging Financial Market	14	0.88	866	54.12
118	Kim Go, 2006, Small Bus Econ	Do Equally Owned Small Businesses Have Equal Access To Credit?	0	0.00	18	1.20
119	Krasniqi Ba, 2010, Int Entrep Manag J	Are Small Firms Really Credit Constrained? Empirical Evidence From Kosova	0	0.00	33	3.00
120	Kumar A, 2019, Econ Model	Usage Of Formal Financial Services In India: Demand Barriers Or Supply Constraints?	0	0.00	1	0.50
121	Kumar A, 2020, J Agr Resour Econ	Access To Credit And Economic Well-Being Of Rural Households: Evidence From Eastern India	0	0.00	5	5.00
122	Kumar Sm, 2013, World Dev	Does Access To Formal Agricultural Credit Depend On Caste?	0	0.00	26	3.25
123	Kwong C, 2012, Int J Entrep Behav R	Differences In Perceptions Of Access To Finance Between Potential Male And Female Entrepreneurs: Evidence From The Uk	1	0.11	36	4.00
124	Ky Ss, 2021, Telecommun Policy	Friends Or Foes? Mobile Money Interaction With Formal And Informal Finance	0		1	Inf
125	Lacour-Little M, 1998, J Real Estate Financ	Are Minorities Or Minority Neighborhoods More Likely To Get Low Appraisals?	0	0.00	11	0.48
126	Ladd Hf, 1998, J Econ Perspect	Evidence On Discrimination In Mortgage Lending	15	0.65	156	6.78
127	Lee J, 2013, Asian Women	Women's Access To Credit: Asian Women's Double Burden	0	0.00	2	0.25
128	Li Ly, 2020, Econ Res-Ekon Istraz	Political Connections And Household Access To Bank Loans: Evidence From China	0	0.00	2	2.00
129	Lin Lq, 2019, Sustainability-Basel	Credit Constraints On Farm Household Welfare In Rural China: Evidence From Fujian Province	0	0.00	2	1.00
130	Lin Lq, 2019, Sustainability-Basel-A	Rural Credit Constraint And Informal Rural Credit Accessibility In China	2	1.00	16	8.00
131	Linh Tn, 2019, Sustainability-Basel	Access To Rural Credit Markets In Developing Countries, The Case Of Vietnam: A Literature Review	1	0.50	13	6.50
132	Lotto J, 2018, Sustainability-Basel	Examination Of The Status Of Financial Inclusion And Its Determinants In Tanzania	0	0.00	7	2.33
133	Lu Zf, 2012, J Bank Financ	Bank Discrimination, Holding Bank Ownership, And Economic Consequences: Evidence From China	0	0.00	54	6.00
134	Luo Dl, 2014, Emerg Mark Financ Tr	Political Connections And Bank Lines Of Credit	0	0.00	21	3.00
135	Malapit Hjl, 2012, Fem Econ	Are Women More Likely To Be Credit Constrained? Evidence From Low-Income Urban Households In The Philippines	2	0.22	8	0.89
136	Malesky Ej, 2009, J Law Econ Organ	Where Is Credit Due? Legal Institutions, Connections, And The Efficiency Of Bank Lending In Vietnam	0	0.00	49	4.08
137	Marlow S, 2005, Entrep Theory Pract	All Credit To Men? Entrepreneurship, Finance, And Gender	4	0.25	334	20.88
138	Marshall Jn, 2004, Environ Plann A	Financial Institutions In Disadvantaged Areas: A Comparative Analysis Of Policies Encouraging Financial Inclusion In Britain And The United States	0	0.00	51	3.00
139	Martinez Lb, 2020, Acad-Rev Latinoam Ad	Evolution Of Financial Inclusion In Latin America Management Area: Business Economics	0	0.00	0	0.00

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
140	Mason Dr, 2014, J Int Dev	Who Gets What? Determinants Of Loan Size And Credit Rationing Among Microcredit Borrowers: Evidence From Nicaragua	0	0.00	9	1.29
141	Mindra R, 2017, Int J Bank Mark	Financial Self-Efficacy: A Determinant Of Financial Inclusion	0	0.00	10	2.50
142	Mndolwa Fd, 2020, Afr Dev Rev	Gender Disparities In Financial Inclusion: Insights From Tanzania	0	0.00	0	0.00
143	Moahid M, 2020, Sustainability-Basel	Factors Affecting Farmers' Access To Formal And Informal Credit: Evidence From Rural Afghanistan	1	1.00	8	8.00
144	Mocan Hn, 1997, Appl Econ	Credit Rationing, Deregulation And Race In Mortgage Lending In The United States: 1960-90	0	0.00	0	0.00
145	Morgan Pj, 2020, J Asian Econ	Financial Literacy, Financial Inclusion, And Savings Behavior In Laos	0	0.00	5	5.00
146	Moro A, 2017, J Bank Financ	Does A Manager's Gender Matter When Accessing Credit? Evidence From European Data	3	0.75	25	6.25
147	Morsy H, 2020, Q Rev Econ Financ	Access To Finance - Mind The Gender Gap	1	1.00	5	5.00
148	Mpuga P, 2010, Afr Dev Rev	Constraints In Access To And Demand For Rural Credit: Evidence From Uganda	0	0.00	28	2.55
149	Muravyev A, 2009, J Comp Econ	Entrepreneurs' Gender And Financial Constraints: Evidence From International Data	20	1.67	145	12.08
150	Mushtaq R, 2019, Technol Soc	Microfinance, Financial Inclusion And Ict: Implications For Poverty And Inequality	0	0.00	21	10.50
151	Myers Sl, 1995, Soc Sci Quart	Racial-Discrimination In Housing Markets - Accounting For Credit Risk	2	0.08	27	1.04
152	Neaime S, 2018, Financ Res Lett	Financial Inclusion And Stability In Mena: Evidence From Poverty And Inequality	1	0.33	50	16.67
153	Nguyen Lt, 2019, Appl Econ	Sme Credit Constraints In Asia's Rising Economic Star: Fresh Empirical Evidence From Vietnam	0	0.00	4	2.00
154	Nothaft Fe, 2002, J Hous Econ	Do Mortgage Rates Vary By Neighborhood? Implications For Loan Pricing And Redlining	2	0.11	15	0.79
155	Okten C, 2004, World Dev	Social Networks And Credit Access In Indonesia	5	0.29	80	4.71
156	Okurut Fn, 2005, S Afr J Econ	Credit Demand And Credit Rationing In The Informal Financial Sector In Uganda	3	0.19	14	0.88
157	Okurut Fn, 2007, S Afr J Econ Manag S	Credit Market Access In Uganda: Evidence From Household Survey Data 1999/2000	0	0.00	4	0.29
158	Olney Ml, 1998, J Econ Hist	When Your Word Is Not Enough: Race, Collateral, And Household Credit	1	0.04	15	0.65
159	Oluwasola O, 2008, Outlook Agr	Determinants Of Agricultural Credit Demand And Supply Among Small-Scale Farmers In Nigeria	0	0.00	8	0.62
160	Ongena S, 2016, J Money Credit Bank	Gender Bias And Credit Access	1	0.20	16	3.20
161	Orser Bj, 2006, Entrep Theory Pract	Women Entrepreneurs And Financial Capital	3	0.20	145	9.67
162	Osei-Tutu F, 2021, Econ Transit I Chang	Sex, Language And Financial Inclusion*	0		0	
163	Owen Al, 2018, Rev Dev Financ	Bank Concentration, Competition, And Financial Inclusion	1	0.33	29	9.67
164	Ozili Pk, 2018, Borsa Istanb Rev	Impact Of Digital Finance On Financial Inclusion And Stability	0	0.00	67	22.33

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
165	Pager D, 2008, Annu Rev Sociol	The Sociology Of Discrimination: Racial Discrimination In Employment, Housing, Credit, And Consumer Markets	4	0.31	689	53.00
166	Park Cy, 2018, Singap Econ Rev	Financial Inclusion, Poverty, And Income Inequality	0	0.00	27	9.00
167	Pellegrina Ld, 2011, World Dev	Microfinance And Investment: A Comparison With Bank And Informal Lending	0	0.00	23	2.30
168	Peng Hf, 2017, Account Financ	Political Connections Of The Board Of Directors And Credit Financing: Evidence From Chinese Private Enterprises	0	0.00	20	5.00
169	Pham T, 2018, World Dev	Discrimination, Social Capital, And Financial Constraints: The Case Of Viet Nam	0	0.00	28	9.33
170	Pham Ttt, 2007, Econ Transit	Lending Policies Of Informal, Formal And Semiformal Lenders - Evidence From Vietnam	2	0.14	20	1.43
171	Pitt Mm, 1998, J Polit Econ	The Impact Of Group-Based Credit Programs On Poor Households In Bangladesh: Does The Gender Of Participants Matter?	13	0.57	500	21.74
172	Presbitero Af, 2014, J Dev Stud	Barking Up The Wrong Tree? Measuring Gender Gaps In Firm's Access To Finance	3	0.43	15	2.14
173	Ramachandran V, 1999, J Dev Stud	Minority Entrepreneurs And Firm Performance In Sub-Saharan Africa	1	0.05	51	2.32
174	Rand J, 2007, Small Bus Econ	Credit Constraints And Determinants Of The Cost Of Capital In Vietnamese Manufacturing	3	0.21	30	2.14
175	Rao S, 2015, J Fam Econ Iss	Gender, Household Structure And Financial Participation In The United States	0	0.00	7	1.17
176	Reyes A, 2011, J Dev Stud	The Credit Constraints Of Market-Oriented Farmers In Chile	0	0.00	14	1.40
177	Ruziev K, 2015, J Dev Stud	Connectedness And Sme Financing In Post-Communist Economies: Evidence From Uzbekistan	1	0.17	7	1.17
178	Ruziev K, 2019, Post-Communist Econ	Does Connectedness Improve Smes' Access To Formal Finance? Evidence From Post-Communist Economies	0	0.00	2	1.00
179	Sarma M, 2011, J Int Dev	Financial Inclusion And Development	5	0.50	170	17.00
180	Shihadeh Fh, 2018, Int J Islamic Middle	How Individual's Characteristics Influence Financial Inclusion: Evidence From Menap	1	0.33	9	3.00
181	Shoji M, 2012, World Dev	Social Capital Formation And Credit Access: Evidence From Sri Lanka	3	0.33	29	3.22
182	Smallbone D, 2003, Int Small Bus J	Access To Finance By Ethnic Minority Businesses In The Uk	2	0.11	55	3.06
183	Solo Tm, 2008, Environ Urban	Financial Exclusion In Latin America - Or The Social Costs Of Not Banking The Urban Poor	0	0.00	23	1.77
184	Squires Gd, 1995, Soc Sci Quart	Does Anybody Who Works Here Look Like Me - Mortgage Lending, Race, And Lender Employment	0	0.00	26	1.00
185	Squires Gd, 1996, Environ Plann A	Mortgage Lending And Race: Is Discrimination Still A Factor?	0	0.00	8	0.32
186	Stefani Ml, 2015, Cesifo Econ Stud	Small Firms' Credit Access In The Euro Area: Does Gender Matter?	0	0.00	11	1.83
187	Steil Jp, 2018, Housing Stud	The Social Structure Of Mortgage Discrimination	0	0.00	11	3.67
188	Storey Dj, 2004, Small Bus Econ	Racial And Gender Discrimination In The Micro Firms Credit Market? Evidence From Trinidad And Tobago	10	0.59	58	3.41

	1st Author, Year, Journal	Title	TLC	TLC/t	TGC	TGC/t
189	Swaminathan H, 2010, World Dev	Impact Of Access To Credit On Labor Allocation Patterns In Malawi	1	0.09	22	2.00
190	Swamy V, 2014, World Dev	Financial Inclusion, Gender Dimension, And Economic Impact On Poor Households	5	0.71	70	10.00
191	Talavera O, 2012, Emerg Mark Financ Tr	Social Capital And Access To Bank Financing: The Case Of Chinese Entrepreneurs	2	0.22	32	3.56
192	Tambunlertchai K, 2018, Singap Econ Rev	Determinants And Barriers To Financial Inclusion In Myanmar: What Determines Access To Financial Services And What Hinders It?	0	0.00	1	0.33
193	Tran Tkv, 2018, Asian-Pac Econ Lit	Gender Differences In Formal Credit Approaches: Rural Households In Vietnam	0	0.00	5	1.67
194	Ullah A, 2020, Agriculture-Basel	Factors Determining Farmers' Access To And Sources Of Credit: Evidence From The Rain-Fed Zone Of Pakistan	0	0.00	2	2.00
195	Ullah B, 2020, Q Rev Econ Financ	Financial Constraints, Corruption, And Sme Growth In Transition Economies	0	0.00	17	17.00
196	Wang Xh, 2017, Appl Econ	Financial Inclusion: Measurement, Spatial Effects And Influencing Factors	1	0.25	28	7.00
197	Wellalage N, 2017, Res Int Bus Financ	Access To Credit By Smes In South Asia: Do Women Entrepreneurs Face Discrimination	1	0.25	14	3.50
198	Wellalage Nh, 2019, J Bus Ethics	Corruption, Gender And Credit Constraints: Evidence From South Asian Smes	0	0.00	21	10.50
199	Wheeler Ch, 2015, J Hous Econ	Racial Differences In Mortgage Denials Over The Housing Cycle: Evidence From Us Metropolitan Areas	1	0.17	2	0.33
200	Witbooi M, 2011, Afr J Bus Manage	Indigenous Female Entrepreneurship: Analytical Study On Access To Finance For Women Entrepreneurs In South Africa	0	0.00	13	1.30
201	Wyly Ek, 2002, Econ Geogr	The Disappearance Of Race In Mortgage Lending	3	0.16	15	0.79
202	Wyly Ek, 2004, Environ Plann A	Gentrification, Segregation, And Discrimination In The American Urban System	0	0.00	105	6.18
203	Wyly Ek, 2007, Environ Plann A	Race, Gender, And Statistical Representation: Predatory Mortgage Lending And The Us Community Reinvestment Movement	1	0.07	28	2.00
204	Xu Nn, 2020, Financ Res Lett	Financial Literacy And Formal Credit Accessibility: Evidence From Informal Businesses In China	0	0.00	4	4.00
205	Xu Xy, 2020, Financ Res Lett	Trust And Financial Inclusion: A Cross-Country Study	0	0.00	5	5.00
206	Yeh Yh, 2013, Pac-Basin Financ J	Political Connections, Corporate Governance And Preferential Bank Loans	0	0.00	41	5.12
207	Yuan Y, 2015, China Econ Rev	Are Poor Able To Access The Informal Credit Market? Evidence From Rural Households In China	3	0.50	28	4.67
208	Zhao Jm, 2014, China Agr Econ Rev	Effects Of Credit Constraints On Rural Household Technical Efficiency Evidence From A City In Northern China	2	0.29	12	1.71
209	Zins A, 2016, Rev Dev Financ	The Determinants Of Financial Inclusion In Africa	7	1.40	89	17.80
210	Zonta Mm, 2012, Environ Plann A	Applying For Home Mortgages In Immigrant Communities: The Case Of Asian Applicants In Los Angeles	0	0.00	2	0.22

A.2 Journals and Affiliations

Table A.2: Journals with Author Universities

	Journal	Article.No.	Author Affiliation
1	World Development	14	Bilkent Univ (1);Bocconi Univ (1);German Inst Econ Res Diw Berlin (1);Humboldt Univ (1);Indian Inst Management Bangalore (1);Indiana Univ Purdue Univ (1);Jica Res Inst (1);Leibniz Univ Hannover (2);Ohio State Univ (1);Penn State Univ (1);Reserve Bank India (1);Ryukoku Univ (1);Seijo Univ (1);Swansea Univ (1);Tilburg Univ (1);Univ E Anglia (1);Univ Libre Bruxelles (1);Univ Milano Bicocca (1);Univ Tokyo (1);Univ Washington (2);World Bank (1);NA (1)
2	Journal Of Development Studies	10	Aalto Univ (1);Aberystwyth Univ (1);Ctr European Res Microfinance Cermi (1);Int Monetary Fund (1);Sapienza Univ Rome (1);Univ Bath (2);Univ Calif Davis (1);Univ Copenhagen (2);Univ Groningen (1);Univ Liverpool (1);Univ Namur (1);Univ Paris Diderot (1);Univ Pavia (1);Univ Santo Tomas (1);Univ Tampere (1);Univ Vaasa (1);Univ W England (1);Univ Washington (1);Wageningen Univ (1);World Bank (1)
3	Journal Of Banking & Finance	8	Asian Dev Bank Inst (1);Ca Foscari Univ Venice (1);Chinese Univ Hong Kong (1);City Univ Hong Kong (1);City Univ London (1);Cranfield Univ (1);Ctr Econ Policy Res (1);Fed Reserve Bank Atlanta (1);Getulio Vargas Fdn (1);Jinan Univ (1);Lingnan Univ (1);Natl Univ Singapore (1);Peking Univ (1);Stockholm Univ (1);Univ Int Business And Econ (1);Univ Leicester (1);Univ Sci And Technol China (1);Univ Tennessee (1);Wissensch Zentrum Berlin Sozialforsch (1);World Bank (1);Zentrum Europaische Wirtschaftsforsch (1)
4	Small Business Economics	7	Bindura Univ Sci Educ (1);Ctr European Res Microfinance Cermi (1);Indian Inst Technol Madras (1);Off President (1);Rgm Sch Intelligence (1);Sikkim Univ (1);Tech Univ Cartagena (1);Univ Copenhagen (1);Univ Leeds (1);Univ Libre Bruxelles (2);Univ Michigan (1);Univ Murcia (1);Univ Warwick (1)
5	Environment And Planning A	5	Illinois State Univ (1);Univ British Columbia (2);Univ Newcastle Upon Tyne (1);Univ Washington (1);Univ Wisconsin (1);Virginia Commonwealth Univ (1)
6	Journal Of International Development	5	Birmingham City Univ (1);Hamilton Coll (1);Inst Studies Ind Dev (1);Jawaharlal Nehru Univ (1);Northumbria Univ (1);Pan Atlantic Univ (2);Univ Calif Los Angeles (1);Univ Wolverhampton (1)
7	Sustainability	5	Dalian Minzu Univ (2);Fujian Agr And Forestry Univ (2);Hiroshima Univ (1);Hue Univ (1);Inst Finance Management (1);Lincoln Univ (3);Nangarhar Univ (1);Natl Econ Univ (1);Univ Danang (2);Univ Liege (1)
8	Emerging Markets Finance And Trade	4	Kadir Has Univ (1);Robert Gordon Univ (1);Sichuan Agr Univ (1);Sichuan Univ (1);Sun Yat Sen Univ (1);Tianjin Univ (1);Univ E Anglia (1)
9	Entrepreneurship Theory And Practice	4	Babson Coll (1);Carleton Univ (1);Northeastern Univ (3);Univ Cent Lancashire (1);Univ Glasgow (1);Univ Ottawa (1);Univ Sterling (1);Univ Strathclyde (1);NA (1)
10	Finance Research Letters	4	Amer Univ Beirut (1);San Jose State Univ (2);Southwestern Univ Finance And Econ (1);Univ Illinois (1);Univ Perugia (1);Zhongnan Univ Econ And Law (1)
11	Journal Of Real Estate Finance And Economics	4	Fed Reserve Board (1);Thema Univ Cergy Pontoise (1);Univ Georgia (1);Univ So Calif (1);Univ Wisconsin (1)
12	Journal Of Urban Economics	4	Duke Univ (1);Georgia State Univ (1);Louisiana State Univ (1);Marquette Univ (1);New Englannd Publ Policy Ctr (1);Syracuse Univ (3);Texas Christian Univ (1);Univ Illinois (1);Univ Penn (1);Univ So Calif (2);Univ Toronto (1)

	Journal	Article.No.	Author Affiliation
13	Applied Economics	3	Griffith Univ (1);Hunan Univ (1);Natl Bur Econ Res (1)
14	Journal Of Business	3	Fed Reserve Syst (1);Georgetown Univ (1);Hitotsubashi Univ (1);Univ Arkansas (1);NA (1)
15	Journal Of Money Credit And Banking	3	Ctr Naval Anal (1);Fed Reserve Board Governors (1);Georgetown Univ (1);Univ Zurich (1)
16	African Development Review-Revue Africaine De Developpement	2	Univ Cape Town (1);NA (1)
17	American Economic Review	2	Univ Johannesburg (1);Univ Kansas (3);Univ Massachusetts (1);Wheaton Coll (1)
18	China Economic Review	2	Bank Finland Inst Econ Transit Bofit (1);Charles Univ Prague (1);Univ Strasbourg (1);NA (1)
19	Feminist Economics	2	Trident Univ Int (1);Univ Calif Davis (1);Univ Calif Riverside (1);Univ Leeds (1);World Bank (1)
20	Housing Policy Debate	2	Fed Reserve Syst (1);Univ Nevada (1);Univ Penn (1);Univ Tennessee (1)
21	International Journal Of Entrepreneurial Behavior & Research	2	Cardiff Metropolitan Univ (1);Univ Essex (1);Univ Rostock (1);Univ Wales Coll Cardiff (1)
22	International Small Business Journal-Researching Entrepreneurship	2	Univ Birmingham (3);Univ Strathclyde (2);Univ Warwick (1)
23	Journal Of African Economies	2	African Econ Res Consortium (1);Bocconi Univ (1);Boston Coll (1);City Univ London (1);Shanghai Jiao Tong Univ (1);Tilburg Univ (1);Univ Chile (1);Univ London Imperial Coll Sci Technol And Med (1);Univ Maryland (1);World Bank (2)
24	Journal Of Asian Economics	2	Asian Dev Bank Inst (1);Res Inst Econ Trade And Ind Rieti (1);Ton Duc Thang Univ (1);Waseda Univ (1)
25	Journal Of Business Ethics	2	Ctr Emile Bernheim (1);Montpellier Business Sch (1);Univ Waikato (1)
26	Journal Of Comparative Economics	2	Free Univ Berlin (1);German Inst Econ Res Diw Berlin (1);Inst Study Labor (1);Kyiv Sch Econ (1);Robert Gordon Univ (1);St Petersburg State Univ (1);Univ Amsterdam (1)
27	Journal Of Corporate Finance	2	Cheung Kong Grad Sch Business (1);Directorate Gen Econ Stat And Res (1);Peking Univ (1);World Bank (1)
28	Journal Of Family And Economic Issues	2	Brandeis Univ (2);Int Food Policy Res Inst (1);Louisiana State Univ (1);Silver Sch Social Work (1)
29	Journal Of Finance	2	Harvard Univ (1);Univ Houston (1)
30	Journal Of Financial Economics	2	Int Monetary Fund (1);Tilburg Univ (1);Univ Amsterdam (1);Univ Maryland (1);World Bank (2)
31	Journal Of Financial Services Research	2	Bangor Univ (1);Loughborough Univ Technol (1);Natl Taichung Univ Sci And Technol (1);Natl Taiwan Univ (1);Natl Yunlin Univ Sci And Technol (1)
32	Journal Of Financial Stability	2	Brock Univ (1);Depaul Univ (1);Nankai Univ (1);Univ Essex (1);Univ Glasgow (1);Univ Leeds (1)
33	Journal Of Housing Economics	2	George Washington Univ (1);NA (1)
34	Journal Of Small Business Management	2	Duke Univ (2);Univ Hartford (1)
35	Quarterly Review Of Economics And Finance	2	Macroecon Policy And Res (1);Morgan State Univ (1)

	Journal	Article.No.	Author Affiliation
36	Regional Science And Urban Economics	2	Arizona State Univ (1);Fed Reserve Bank St Louis (1);George Washington Univ (1)
37	Research In International Business And Finance	2	Sch Business (1);Univ Ghana (1);Waikato Management Sch (1)
38	Review Of Black Political Economy	2	Coll Wooster (1);Urban Inst (1);NA (1)
39	Review Of Development Finance	2	Hamilton Coll (1);Univ Strasbourg (2)
40	Review Of Economics And Statistics	2	Dartmouth Coll (1);Univ Chicago (1);Univ Michigan (1);Wellesley Coll (1)
41	Singapore Economic Review	2	Chulalongkorn Univ (1);Ortigas Ctr Mandaluyong (1);Seacen Ctr (1)
42	Social Science Quarterly	2	Univ Maryland (1);Univ Wisconsin (1)
43	Academia-Revista Latinoamericana De Administracion	1	Univ Nacl Sur (1);Univ Prov Sudoeste (1)
44	Accounting And Finance	1	Shandong Univ Finance And Econ (1);Tsinghua Univ (1)
45	African And Asian Studies	1	Univ Antwerp (2);Univ Ghent (1)
46	African Journal Of Business Management	1	Cape Peninsula Univ Technol (1);Univ Johannesburg (1)
47	Agricultural Economics-Zemedelska Ekonomika	1	Agr Univ Krakow (1);Slovak Univ Agr (1)
48	Agriculture-Basel	1	Eberswalde Univ Sustainable Dev (1);Leibniz Ctr Agr Landscape Res Zalf (1);Pmas Arid Agr Univ (1);Univ Agr (1);Univ Alberta (1)
49	Annual Review Of Economics, Vol 5	1	Brown Univ (1);Harvard Univ (1);Natl Bur Econ Res (1);Us Mil Acad (1)
50	Annual Review Of Sociology	1	Princeton Univ (1)
51	Asian-Pacific Economic Literature	1	Cent Inst Econ Management (1);China Agr Univ (1);Comsats Inst Informat Technol (1);Hanoi Univ (1);Nanjing Univ Informat Sci And Technol (1)
52	Asian Women	1	Korea Univ (1)
53	Borsa Istanbul Review	1	Univ Essex (1)
54	Brookings Papers On Economic Activity	1	World Bank (1)
55	Cesifo Economic Studies	1	Econ Res Unit (2)
56	China Agricultural Economic Review	1	Cent Univ Finance And Econ (1);Univ Illinois (1)
57	Developing Economies	1	Univ Wageningen And Res Ctr (1)
58	Economic Development Quarterly	1	Ewing Marion Kauffman Fdn (1);Wayne State Univ (1)

	Journal	Article.No.	Author Affiliation
59	Economic Geography	1	Rutgers State Univ (2);Univ Georgia (1)
60	Economic Modelling	1	Indian Inst Technol (1)
61	Economic Policy	1	Goethe Univ Frankfurt (1);Swiss Natl Bank (1);Tilburg Univ (1)
62	Economic Research-Ekonomska Istrazivanja	1	Northwest Univ (1);Univ Groningen (1)
63	Economics Letters	1	Univ Copenhagen (2)
64	Economics Of Transition	1	Univ Groningen (1)
65	Economics Of Transition And Institutional Change	1	Mgimo Univ (1);Univ Strasbourg (2)
66	Environment And Planning C-Government And Policy	1	Univ Stellenbosch (1)
67	Environment And Urbanization	1	World Bank (1)
68	European Financial Management	1	Sapienza Univ Rome (1);Univ Leeds (1);Univ Trieste (1)
69	European Journal Of Development Research	1	NA (1)
70	European Journal Of Finance	1	Bangor Univ (1);Univ Hull (1)
71	Housing Studies	1	Brigham Young Univ (1);Northeastern Univ (1);Princeton Univ (1)
72	International Entrepreneurship And Management Journal	1	Staffordshire Univ (1)
73	International Journal Of Bank Marketing	1	Makerere Univ (2);Univ Witwatersrand (2)
74	International Journal Of Emerging Markets	1	Makerere Univ (3)
75	International Journal Of Finance & Economics	1	Univ Waikato (1)
76	International Journal Of Islamic And Middle Eastern Finance And Management	1	Palestine Tech Univ (1)
77	International Small Business Journal	1	De Montfort Univ (1);Middlesex Univ (1);Paisley Business Sch (1)
78	Journal Of Accounting Research	1	City Univ Hong Kong (1);Lingnan Univ (1);Univ Florida (1);Univ Hong Kong (1)
79	Journal Of Agricultural And Resource Economics	1	Arizona State Univ (1);Int Food Policy Res Inst (1)
80	Journal Of Business Venturing	1	Lincoln Univ (1)
81	Journal Of Consumer Affairs	1	Northern Kentucky Univ (1);Univ Kentucky (1);Western Michigan Univ (1)

	Journal	Article.No.	Author Affiliation
82	Journal Of Development Economics	1	Univ Oxford (1)
83	Journal Of Economic Behavior & Organization	1	Natl Council Appl Econ Res (1);Swiss Re Econ Res And Consulting (1);World Bank (1)
84	Journal Of Economic Geography	1	Univ Cambridge (2);Univ Roma Tor Vergata (1)
85	Journal Of Economic History	1	Univ Calif Berkeley (1)
86	Journal Of Economic Issues	1	Jackson State Univ (1)
87	Journal Of Economic Perspectives	1	Duke Univ (1)
88	Journal Of Financial And Quantitative Analysis	1	Cornell Univ (1);Tsinghua Univ (1);Univ Hawaii Manoa (1)
89	Journal Of Financial Intermediation	1	Dev Res Grp (1);Imperial Coll London (1);Univ Penn (1)
90	Journal Of Law Economics & Organization	1	Harvard Univ (1);Univ Calif San Diego (1)
91	Journal Of Policy Modeling	1	Indian Stat Inst (1);Indira Gandhi Inst Dev Res (1)
92	Journal Of Political Economy	1	Brown Univ (1);World Bank (1)
93	Journal Of Real Estate Research	1	NA (1)
94	Journal Of The European Economic Association	1	Harvard Univ (1)
95	Journal Of Urban Affairs	1	Wayne State Univ (1)
96	Kyklos	1	Linnaeus Univ (1)
97	Management Science	1	Catholic Univ Portugal (1);Erasmus Univ (1);Univ Colorado (1);Univ Virginia (1)
98	New England Economic Review	1	Le (Corresponding Author) (1)
99	Outlook On Agriculture	1	Obafemi Awolowo Univ (1)
100	Oxford Bulletin Of Economics And Statistics	1	Peking Univ (1);Texas Aandm Univ (1);World Bank (2);Zhongnan Univ Econ And Law (1)
101	Oxford Economic Papers-New Series	1	Univ Reading (1)
102	Pacific-Basin Finance Journal	1	Fu Jen Catholic Univ (1);Natl Chiao Tung Univ (1);Natl Taiwan Univ (1)
103	Post-Communist Economies	1	Univ West England (1)
104	Quarterly Journal Of Economics	1	Harvard Univ (1);Univ Chicago (1)
105	Real Estate Economics	1	Penn State Univ (1);Univ Georgia (1)
106	Review Of Economic Studies	1	NA (1)
107	Review Of Financial Studies	1	Duke Univ (1);Univ Connecticut (1);Univ Penn (1)

	Journal	Article.No.	Author Affiliation
108	Social Indicators Research	1	Bahcesehir Univ (1);Univ Milano Bicocca (1);World Bank Grp (1)
109	Social Psychology Quarterly	1	Univ Iowa (1)
110	South African Journal Of Economic And Management Sciences	1	Univ Botswana (1);Univ Stellenbosch (1)
111	South African Journal Of Economics	1	Univ Stellenbosch (1)
112	Technology In Society	1	Lyon (1);Paris (1)
113	Telecommunications Policy	1	Univ Limoges (1);Univ Ouahigouya (1)
114	World Bank Economic Review	1	Dev Econ Res Grp (1);Tilburg Univ (1)
115	World Bank Research Observer	1	Gothenburg Univ (1);Univ Oxford (1)

A.3   Biblio Shinyapp

We present a few screenshot of the Biblio Shinyapp to give the reader intuition.
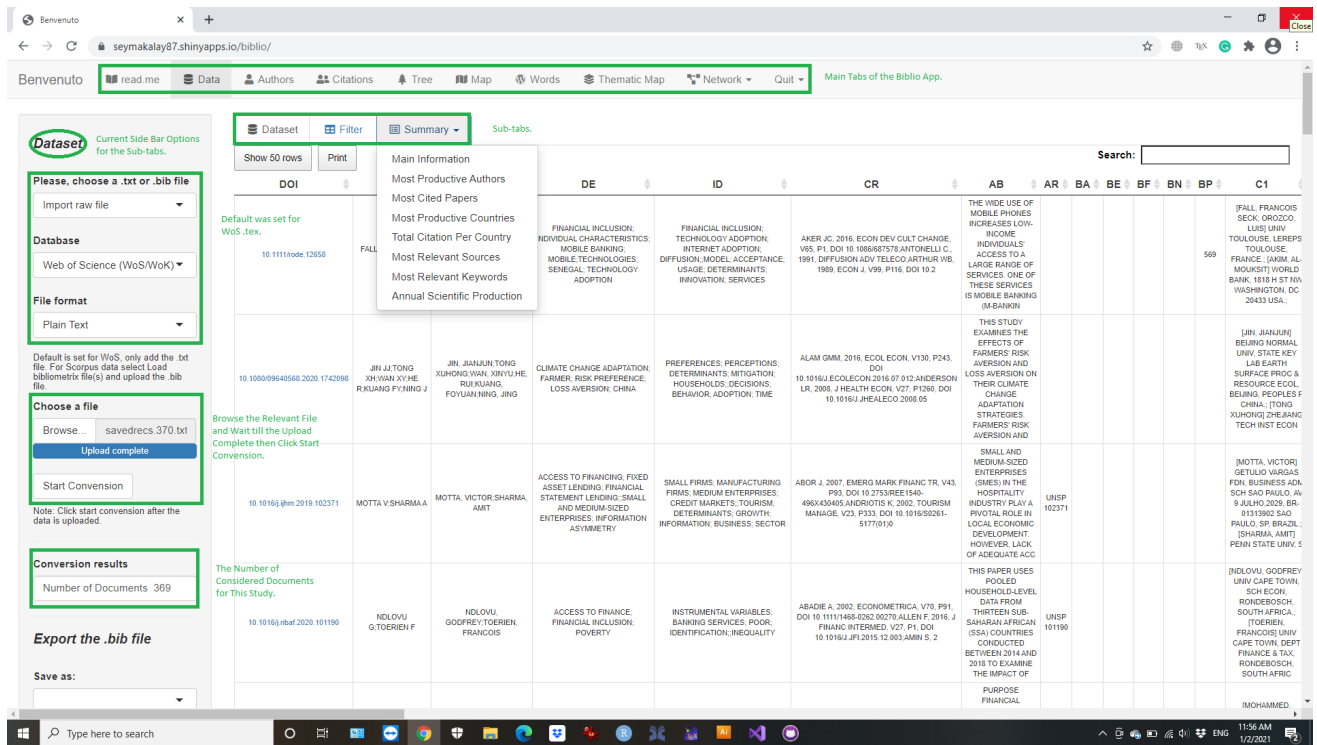
Figure A.1: Data tab.
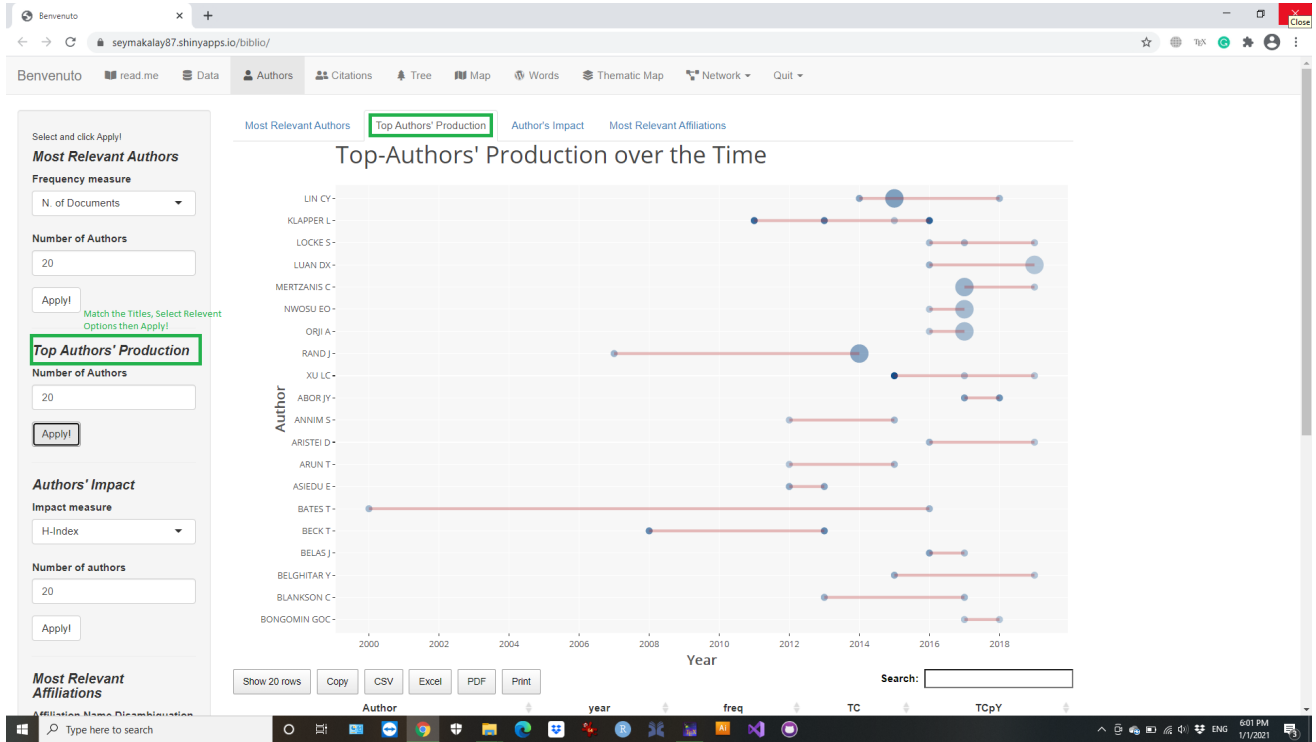
Figure A.2: Authors tab.
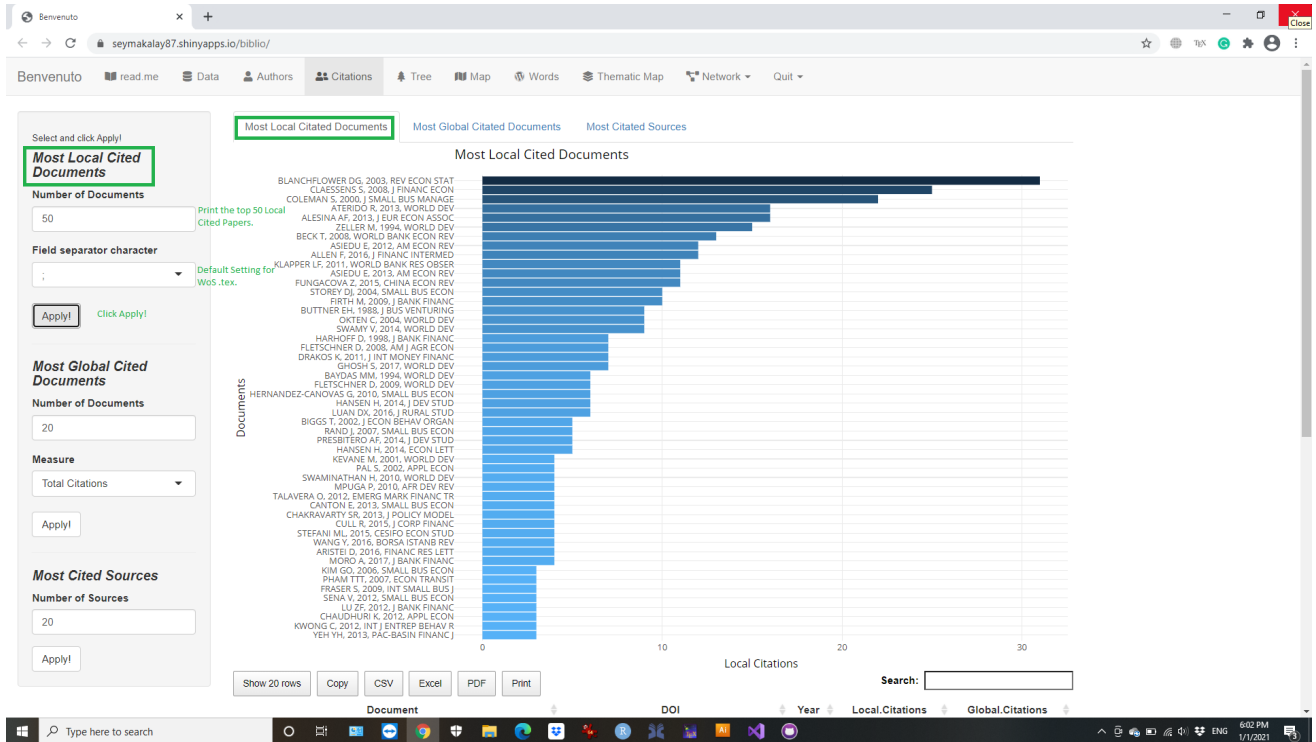
Figure A.3: Citations tab.
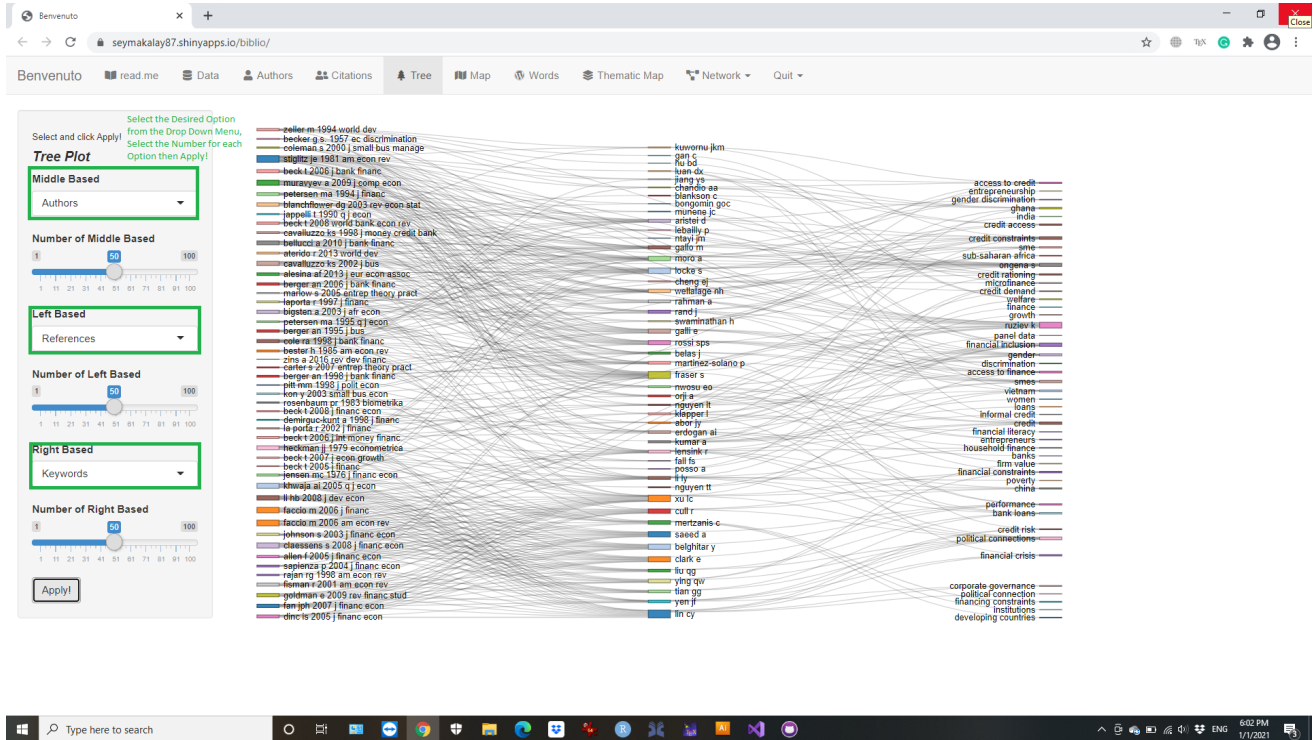
Figure A.4: Tree tab.
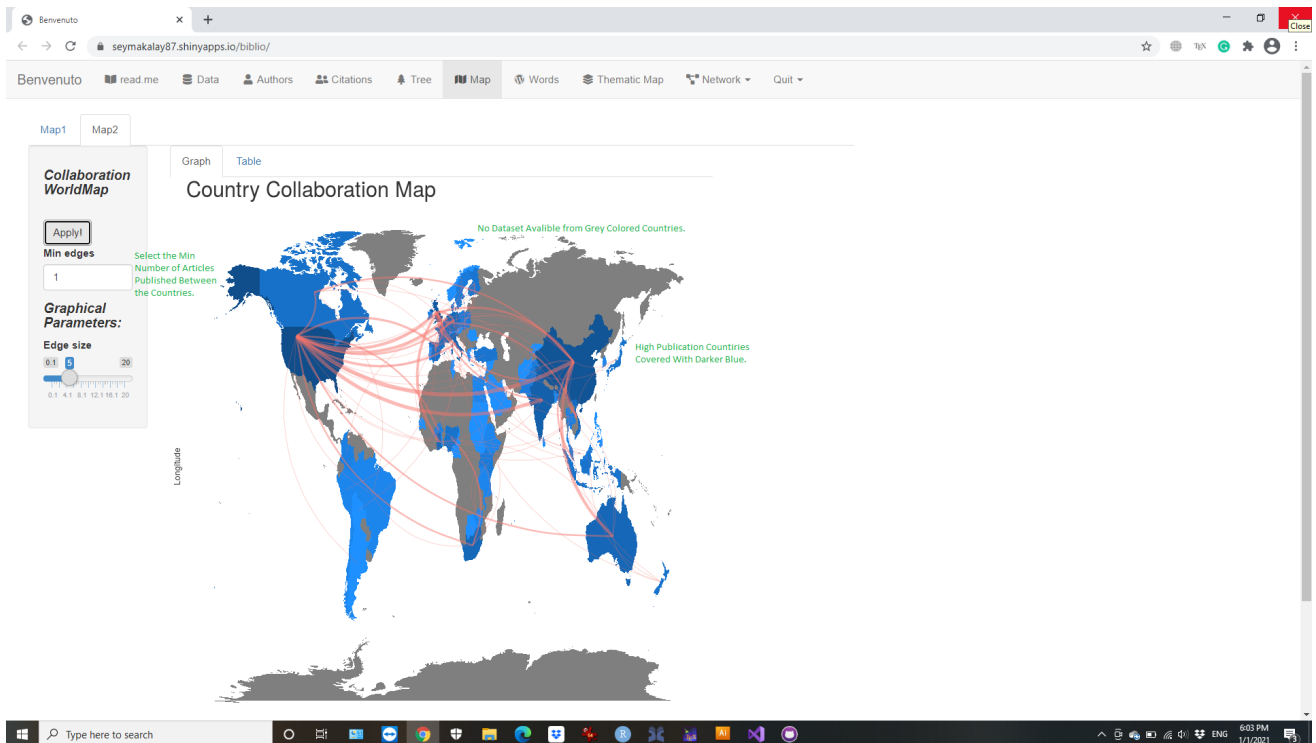
Figure A.5: Map tab.
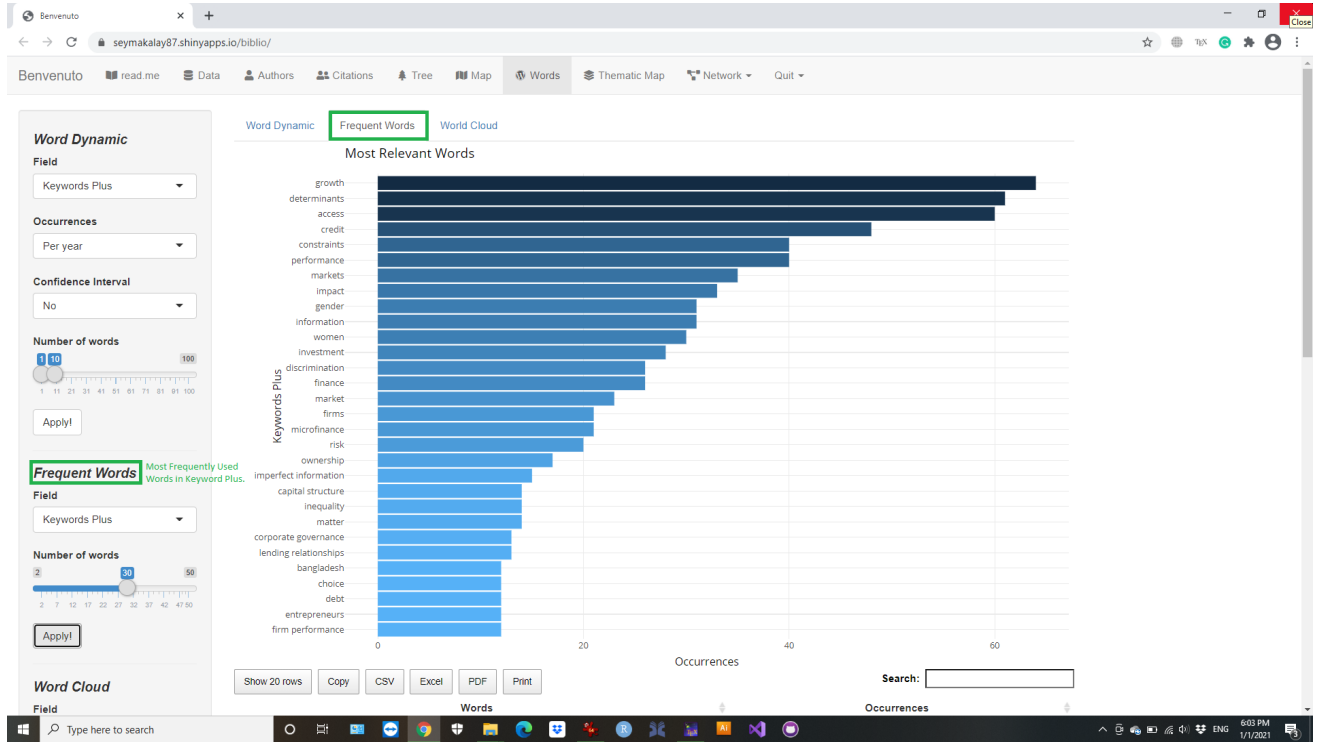
Figure A.6:   Words tab.
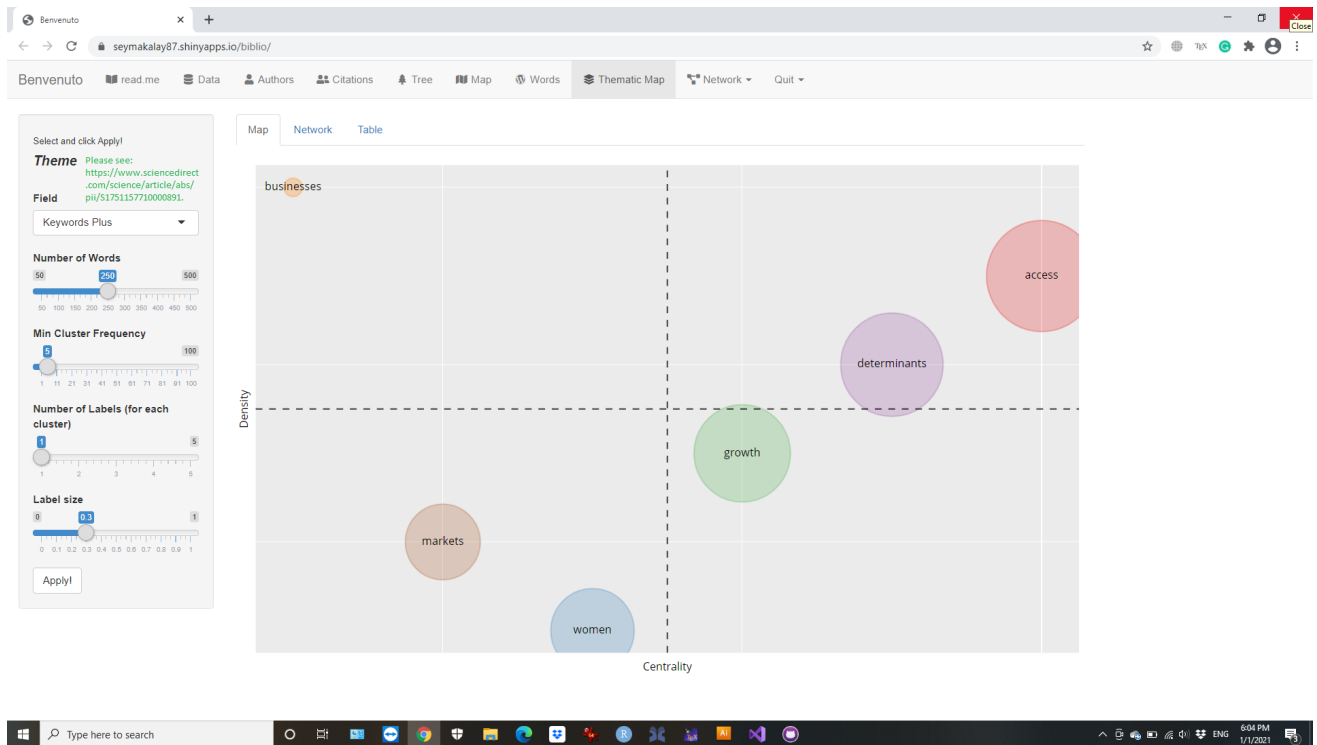
Figure A.7:   Thematic tab.
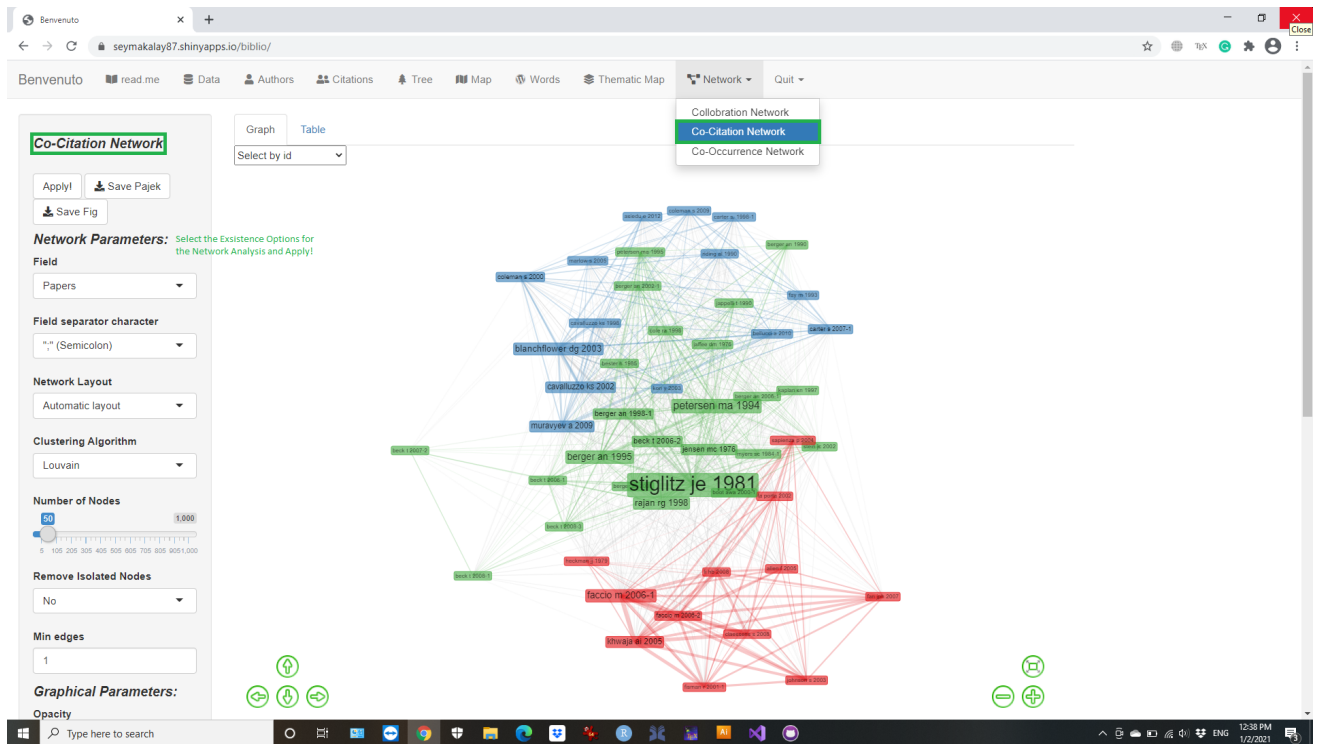
Figure A.8: Network tab.

Appendix B

Appendix of Chapter 3

B.1   Definition of the independent variables

Variable		Definition
independent variables		
x_1	Gender	*Sex.1 = male = 1, Sex.0 = female = 0*
x_2	Marital Status	*married = 1, otherwise = 0*
x_3	Age	household's age, in years.
x_4	Employed	*employed = 1, otherwise = 0*
x_5	Education	*Educ.1 = high school or higher = 1, Educ.0 = otherwise = 0*
x_6	Party	*CCP.1 = affiliation with Chinese Communist Party (CCP) = 1, CCP.0 = otherwise = 0*
x_7	HR	*Urb = urban = 1, Rrl = rural = 0*
x_8	Region	*west, east,* and *center.*
x_9	Fin.Inter	*Fin.Inter = 1 if the household head house is in 1 km range to formal institution, otherwise Fin.Inter = 0*
x_{10}	Fin.Knowledge	*Fin.Knowledge = 1 if the household head has a finance class or defined him/herself having a well financial knowledge, otherwise Fin.Knowledge = 0*
x_{11}	Income	household's income, in CNY.
x_{12}	Net-worth	The value of financial and non-financial assets minus liabilities, in CNY.
x_{13}	NW-HE	Net-worth minus home equity, in CNY.
x_{14}	Liquid Assets	Cash and other easily cash-able assets, in CNY.
dependent variables		
y_1	Access to loan	*Access to loan = 1* if the household head has any type of loan (e.g formal, informal, and/or both); otherwise *Access to loan = 0*.
y_2	Access to loan type	if the household head has formal, informal, or both loans *Access to loan type* is equal to 1, 2, 3, respectively; otherwise *Access to loan type = 0*, which indicates the household head does not have any type of loan.

Note: In equations x_{10}, x_{11}, and x_{12} were use interchangeability.

Table B.1: Definitions of the independent variables. *Note: In equations x_{10}, x_{11}, and x_{12} were use interchangeability.*

B.2   Abbreviations

Abbreviation	Definition
data splits	
Urb & Rrl	CHFS data-set was split into urban and rural.
Educ.0 & Educ.1	CHFS data-set was split into Educ.0 and Educ.1.
CCP.0 & CCP.1	CHFS data-set was split into CCP.0 and CCP.1
Sex.0 & Sex.1	CHFS data-set was split into Sex.0 and Sex.1
linear models	
GLM	To model access to credit. $y = 1$ If the household's head has any type of loan (e.g. Formal, Informal, or Both), otherwise $y = 0$
MLM	To model access to loan type. $y = 1, 2, 3,$ or 4 If the household's head has Formal, Informal, Both, or No.loan, respectively.
ml models	
BAG	Bag tree
RF	Random forest
GBM	Gradiant boosting
loan types	
Fml	If household's head has only Formal loan.
Infm	If household's head has only Informal loan.
Both	If household's head has both Formal and Informal loan.

Note: Urb & Rrl, Educ.0 & Educ.1, CCP.0 & CCP.1, and Sex.0 & Sex.1 are defined in Table B.1.

Table B.2: Definitions of abbreviations.

Appendix C

Appendix of Chapter 4
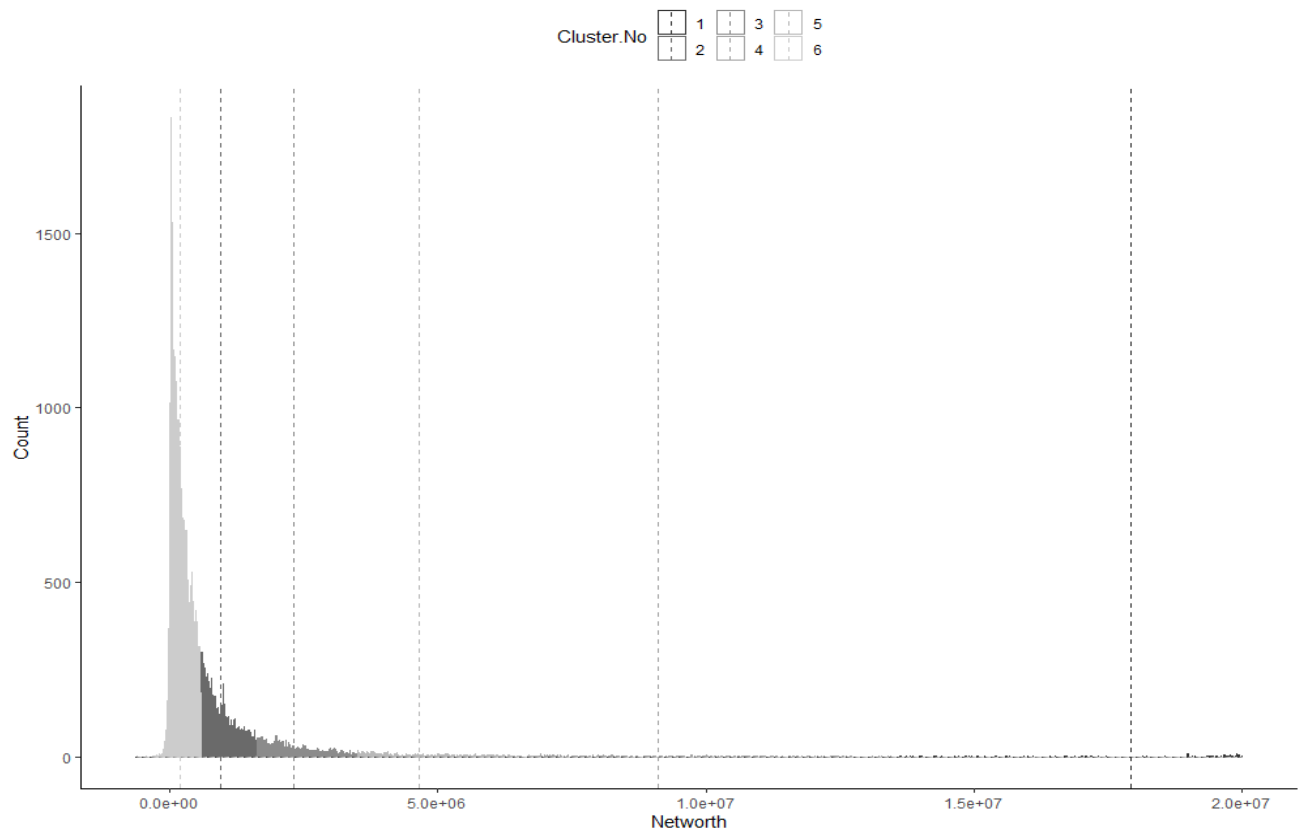
C.1   Cluster Statistics

Figure C.1:   Box plot of continuous variables.

C.2 Micro Shinyapp

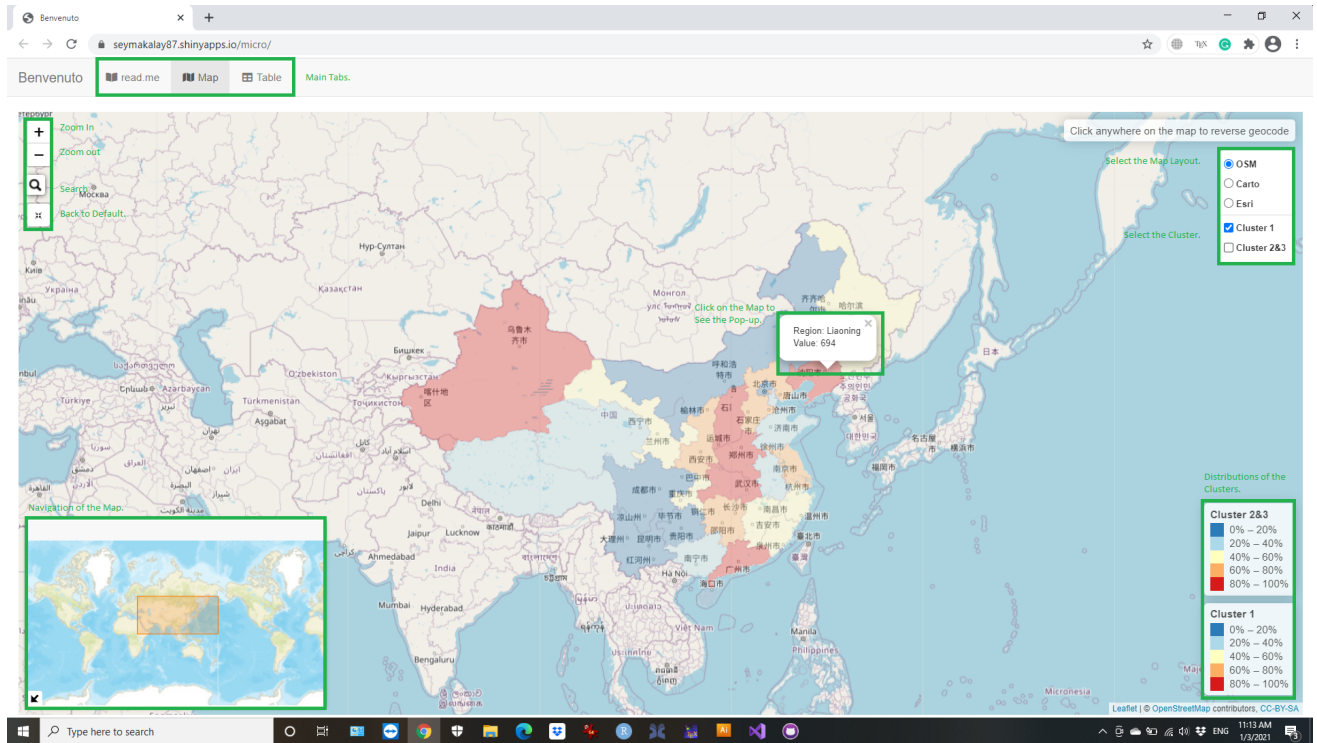We present a few screenshot of the Micro Shinyapp to give the reader intuition.

Figure C.2: Map tab.

Figure C.3:   Table tab.