Essays on Credit Access and Household Finance

Seyma Kalay

Supervisors:
Prof. Michela Cameletti
Prof. Federica Maria Origo

March 6, 2021

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Scienze Aziendali,
Economiche e Metodi Quantitativi

Outline

Outline

Motivation

- 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 government's aim should be toward economically disadvantaged groups to promoting them.

Motivation

- 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 government's aim should be toward economically disadvantaged groups to promoting them.

Motivation

- 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 government's aim should be toward economically disadvantaged groups to promoting them.

Motivation

- 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 government's aim should be toward economically disadvantaged groups to promoting them.

Overview & Contributions

1. Chapter 2
   - reviews the literature on the determinants of finance, using bibliometric techniques, and highlights the influential features of the literature.
2. Chapter 3
   - examines the characteristics of households' head to access to finance for formal, informal, and both sources for different data-splits.
3. Chapter 4
   - clusters the households based on their financial strength and maps the disadvantaged & advanced households' head.

Overview & Contributions

1. Chapter 2
   - reviews the literature on the determinants of finance, using bibliometric techniques, and highlights the influential features of the literature.
2. Chapter 3
   - examines the characteristics of households' head to access to finance for formal, informal, and both sources for different data-splits.
3. Chapter 4
   - clusters the households based on their financial strength and maps the disadvantaged & advanced households' head.

Overview & Contributions

1. Chapter 2
   - reviews the literature on the determinants of finance, using bibliometric techniques, and highlights the influential features of the literature.
2. Chapter 3
   - examines the characteristics of households' head to access to finance for formal, informal, and both sources for different data-splits.
3. Chapter 4
   - clusters the households based on their financial strength and maps the disadvantaged & advanced households' head.

Outline

Motivation

- The amount of academic publications is gradually increasing and it is becoming difficult to identify current trends so scholar uses different quantitative (meta-analysis) and qualitative methods (literature review) to understand earlier findings Aria and Cuccurullo (2017).
- Bibliometric analysis helps to understand the scientific production, intellectual networks, trends and publication patterns between the scholars, institutions and countries Liu (2014), Pinto (2014), Bourdieu (1994), Broadus (1987), Pritchard (1969).
- Bibliometric analysis conducts reproducible literature review concept (Broadus, 1987; Diodato and Gellatly, 2013; Pritchard et al., 1969).

Motivation

- The amount of academic publications is gradually increasing and it is becoming difficult to identify current trends so scholar uses different quantitative (meta-analysis) and qualitative methods (literature review) to understand earlier findings Aria and Cuccurullo (2017).
- Bibliometric analysis helps to understand the scientific production, intellectual networks, trends and publication patterns between the scholars, institutions and countries Liu (2014), Pinto (2014), Bourdieu (1994), Broadus (1987), Pritchard (1969).
- Bibliometric analysis conducts reproducible literature review concept (Broadus, 1987; Diodato and Gellatly, 2013; Pritchard et al., 1969).

Motivation

- The amount of academic publications is gradually increasing and it is becoming difficult to identify current trends so scholar uses different quantitative (meta-analysis) and qualitative methods (literature review) to understand earlier findings Aria and Cuccurullo (2017).
- Bibliometric analysis helps to understand the scientific production, intellectual networks, trends and publication patterns between the scholars, institutions and countries Liu (2014), Pinto (2014), Bourdieu (1994), Broadus (1987), Pritchard (1969).
- Bibliometric analysis conducts reproducible literature review concept (Broadus, 1987; Diodato and Gellatly, 2013; Pritchard et al., 1969).

Workflow

1. **Step 1: Objectives of the Study**
   - What are the influential features, main research streams, and future research questions under "determinants of finance"?
2. **Step 2: Bibliometric Analysis**
   - Study Design: Deciding the research question.
   - Data Collection: Selecting a bibliometric database (e.g., WoS).
     - Articles were selected based on the keyword search which are relevant to "determinants of finance".
     - We only consider published English articles.
     - We read the abstract of each article to decide if "determinants of finance" is not marginally mentioned but have a direct content in the article.
     - Finally, 361 articles were selected to conduct bibliometric analysis.
   - Data Analysis: After exporting .txt file from WoS database, we converted it into a bibliographic dataframe using R and contacted descriptive statistics.
   - Data Visualization: Bibliographic data frame pave the way to network matrix then, network mapping can be conducted.
   - Interpretation: The last stage of the complete bibliometric workflow which helps to researchers make sense of bibliometric's results.
3. **Step 3: Contributions of the Study**
   - We find the influential features of "determinants of finance" literature.
   - We identify two main research streams and we identify 12 future research questions.
   - We provide the code and produce an Shinyapp for the reproducible.

Workflow

1. Step 1: Objectives of the Study
   - What are the influential features, main research streams, and future research questions under "determinants of finance"?
2. Step 2: Bibliometric Analysis
   - Study Design: Deciding the research question.
   - Data Collection: Selecting a bibliometric database (e.g., WoS).
     - Articles were selected based on the keyword search which are relevant to "determinants of finance".
     - We only consider published English articles.
     - We read the abstract of each article to decide if "determinants of finance" is not marginally mentioned but have a direct content in the article.
     - Finally, 361 articles were selected to conduct bibliometric analysis.
   - Data Analysis: After exporting .txt file from WoS database, we converted it into a bibliographic dataframe using R and contacted descriptive statistics.
   - Data Visualization: Bibliographic data frame pave the way to network matrix then, network mapping can be conducted.
   - Interpretation: The last stage of the complete bibliometric workflow which helps to researchers make sense of bibliometric's results.
3. Step 3: Contributions of the Study
   - We find the influential features of "determinants of finance" literature.
   - We identify two main research streams and we identify 12 future research questions.
   - We provide the code and produce an Shinyapp for the reproducible.

Workflow

1. **Step 1: Objectives of the Study**
   - What are the influential features, main research streams, and future research questions under "determinants of finance"?
2. **Step 2: Bibliometric Analysis**
   - Study Design: Deciding the research question.
   - Data Collection: Selecting a bibliometric database (e.g., WoS).
     - Articles were selected based on the keyword search which are relevant to "determinants of finance".
     - We only consider published English articles.
     - We read the abstract of each article to decide if "determinants of finance" is not marginally mentioned but have a direct content in the article.
     - Finally, 361 articles were selected to conduct bibliometric analysis.
   - Data Analysis: After exporting .txt file from WoS database, we converted it into a bibliographic dataframe using R and contacted descriptive statistics.
   - Data Visualization: Bibliographic data frame pave the way to network matrix then, network mapping can be conducted.
   - Interpretation: The last stage of the complete bibliometric workflow which helps to researchers make sense of bibliometric's results.
3. **Step 3: Contributions of the Study**
   - We find the influential features of "determinants of finance" literature.
   - We identify two main research streams and we identify 12 future research questions.
   - We provide the code and produce an Shinyapp for the reproducible.

Bibliometric Coupling

- Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.).
- 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 has the j-th attribute, otherwise $x_{ij} = 0$.
- Let n is the total number of documents (rows) and $\sum x_{+j}$ is 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 total number of attributes which is in the i-th document shown as

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

Bibliometric Coupling

- Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.).
- 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 has the j-th attribute, otherwise $x_{ij} = 0$.
- Let n is the total number of documents (rows) and $\sum x_{+j}$ is 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 total number of attributes which is in the i-th document shown as

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

Bibliometric Coupling

- Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.).
- 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 has the j-th attribute, otherwise $x_{ij} = 0$.
- Let n is the total number of documents (rows) and $\sum x_{+j}$ is 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 total number of attributes which is in the i-th document shown as

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

Bibliometric Coupling

- Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.).
- 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 has the j-th attribute, otherwise $x_{ij} = 0$.
- Let n is the total number of documents (rows) and $\sum x_{+j}$ is 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 total number of attributes which is in the i-th document shown as

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

Bibliometric Coupling

- Documents are connected through a document's attributes (e.g., author, affiliations, publication source, cited references, keywords, etc.).
- 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 has the j-th attribute, otherwise $x_{ij} = 0$.
- Let n is the total number of documents (rows) and $\sum x_{+j}$ is 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 total number of attributes which is in the i-th document shown as

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

Bibliometric Coupling

- 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 (Kessler1963).
- The general formula of bibliometric coupling matrix can be written as

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

- Co-citation, Co-authorship, and Co-word can be presented as

$$B_{co.} = X^T X$$

- Where X is where X is a *Document x Cited reference*, *Document x Author*, and *Document x Word* Matrix and b_{ij} shows the number of total $co - citation$, *collaborations*, and $co -$ *occurrence words* between documents i and j, receptively.

Bibliometric Coupling

- 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 (Kessler1963).
- The general formula of bibliometric coupling matrix can be written as

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

- Co-citation, Co-authorship, and Co-word can be presented as

$$B_{co.} = X^T X$$

- Where X is where X is a *Document x Cited reference*, *Document x Author*, and *Document x Word* Matrix and b_{ij} shows the number of total $co-citation$, *collaborations*, and $co-occurrence$ *words* between documents i and j, receptively.

Bibliometric Coupling

- 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 (Kessler1963).
- The general formula of bibliometric coupling matrix can be written as

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

- Co-citation, Co-authorship, and Co-word can be presented as

$$B_{co.} = X^T X$$

- Where X is where X is a *Document x Cited reference*, *Document x Author*, and *Document x Word* Matrix and b_{ij} shows the number of total $co-citation$, *collaborations*, and $co-occurrence$ *words* between documents i and j, receptively.

Bibliometric Coupling

- 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 (Kessler1963).
- The general formula of bibliometric coupling matrix can be written as

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

- Co-citation, Co-authorship, and Co-word can be presented as

$$B_{co.} = X^T X$$

- Where X is where X is a *Document x Cited reference*, *Document x Author*, and *Document x Word* Matrix and b_{ij} shows the number of total $co - citation$, *collaborations*, and $co -$ *occurrence words* between documents i and j, receptively.

Data Analysis & Visualization

1. Data Analysis
   -
   - Germany, the university of Amsterdam, and Journal of Financial Economics produce more quality papers.
   - Claessens S, Cull R, Firth M, Galiani S, Blanchflower Dg, Fourcade M, Allen F, Harhoff D, Carter S, and Klapper Lf have produced influential papers.
2. Data Visualization
   - The result of co-citation shows that the determinants of finance literature are divided into two main streams: lending to (i) small and (ii) big borrowers.
   - Co-authorship network shows a strong social network among the most productive authors.
   - Co-words network prints out the keywords for each streams.

Data Analysis & Visualization

1. Data Analysis
   -
   - Germany, the university of Amsterdam, and Journal of Financial Economics produce more quality papers.
   - Claessens S, Cull R, Firth M, Galiani S, Blanchflower Dg, Fourcade M, Allen F, Harhoff D, Carter S, and Klapper Lf have produced influential papers.
2. Data Visualization
   - The result of co-citation shows that the determinants of finance literature are divided into two main streams: lending to (i) small and (ii) big borrowers.
   - Co-authorship network shows a strong social network among the most productive authors.
   - Co-words network prints out the keywords for each streams.

Biblio App

- Biblio is a Shinyapp[1], provides an interface for bibliometric analysis, where the results of this study can be reproducible.
- To be able to conduct bibliometric analysis in Shinyapp we need **(i)-*Study design*** and **(ii)-*Data collection***.
- 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.
- The app default data-set 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.

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

Biblio App

- Biblio is a Shinyapp[1], provides an interface for bibliometric analysis, where the results of this study can be reproducible.
- To be able to conduct bibliometric analysis in Shinyapp we need **(i)-*Study design*** and **(ii)-*Data collection***.
- 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.
- The app default data-set 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.

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

Biblio App

- Biblio is a Shinyapp[1], provides an interface for bibliometric analysis, where the results of this study can be reproducible.
- To be able to conduct bibliometric analysis in Shinyapp we need **(i)-*Study design*** and **(ii)-*Data collection***.
- 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.
- The app default data-set 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.

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

Biblio App

- Biblio is a Shinyapp[1], provides an interface for bibliometric analysis, where the results of this study can be reproducible.
- To be able to conduct bibliometric analysis in Shinyapp we need **(i)-*Study design*** and **(ii)-*Data collection***.
- 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.
- The app default data-set 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.

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

Research Questions

Research Stream	Research Direction	Article
(i) small borrowers:		
1	Is it possible to have more fare and equal financial system?	Dwyer2018
2	What kind of policies must target minor groups?	Carter2015
3	What kind of policies should be implemented to ease the women access to finance?	Aterido2013,Ghosh2017
4	Can house owning or land titling programs increase investment on education and helps to access to finance in the long run?	Galiani2010
5	How human capital is affecting approval of loan?	Bruns2008
6	Comparing banking-relationship in develop and developing countries for lending decision making.	UllahBhuiyan2011
7	Why banking-relationship is more important for developing countries?	UllahBhuiyan2011
8	What is the financial accessibility today?	Author's suggestion
9	What is the role of digital finance?	Author's suggestion
10	How access to finance differs between countries where the higher education is free?	Author's suggestion
(ii) big borrowers:		
11	How the stock return or debt reduction differs for the firms, having political ties?	Claessens2008
12	How GDP and welfare of a country are affected negatively, when the companies are favor in access to finance?	Claessens2008

Outline

Motivation

- China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018) and generated 117 million self-employed jobs.
- 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.
- The use of formal financial services is far lower than other emerging (Fung´a˘cov´a et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012).

Motivation

- China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018) and generated 117 million self-employed jobs.
- 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.
- The use of formal financial services is far lower than other emerging (Fung´a˘cov´a et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012).

Motivation

- China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018) and generated 117 million self-employed jobs.
- 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.
- The use of formal financial services is far lower than other emerging (Fung´a˘cov´a et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012).

Motivation

- China is the most populated country in the world with a total population of over 1.4 billion (Demko, 2018) and generated 117 million self-employed jobs.
- 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.
- The use of formal financial services is far lower than other emerging (Fung´aˇcov´a et al., 2014) and high-income economies (Demirguc-Kunt and Klapper, 2012).

Workflow

1. **Step 1: Objectives of the Study**
   - Which data-split can characterize better characteristics to access loan?
2. **Step 2: Methodology** 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 three multi logit and three multi-nominal logit models. Each data-sets were spited into 80:20 train:test groups and 10-CV were applied.
   - Logistic Regression: Access to loan.
     - We have 2 levels ("Access to Loan" = 1, and "Otherwise" = 0) of outcomes.
   - Multi Logit Regression: Access to loan type.
     - We have 4 levels ("Formal Loan" = 1, "Informal Loan" = 2, "Both Loans" = 3, and "Otherwise" = 0) of outcomes.
3. **Step 3: Contributions of the Study**
   - We find the best data-split which can explains better both access to loan and loan type.
   - We find using data-set as a whole shows the lowest predictive power.

Workflow

1. Step 1: Objectives of the Study
   - Which data-split can characterize better characteristics to access loan?
2. Step 2: Methodology 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 three multi logit and three multi-nominal logit models. Each data-sets were spited into 80:20 train:test groups and 10-CV were applied.
   - Logistic Regression: Access to loan.
     - We have 2 levels ("Access to Loan" = 1, and "Otherwise" = 0) of outcomes.
   - Multi Logit Regression: Access to loan type.
     - We have 4 levels ("Formal Loan" = 1, "Informal Loan" = 2, "Both Loans" = 3, and "Otherwise" = 0) of outcomes.
3. Step 3: Contributions of the Study
   - We find the best data-split which can explains better both access to loan and loan type.
   - We find using data-set as a whole shows the lowest predictive power.

Workflow

1. Step 1: Objectives of the Study
   - Which data-split can characterize better characteristics to access loan?
2. Step 2: Methodology 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 three multi logit and three multi-nominal logit models. Each data-sets were spited into 80:20 train:test groups and 10-CV were applied.
   - Logistic Regression: Access to loan.
     - We have 2 levels ("Access to Loan" = 1, and "Otherwise" = 0) of outcomes.
   - Multi Logit Regression: Access to loan type.
     - We have 4 levels ("Formal Loan" = 1, "Informal Loan" = 2, "Both Loans" = 3, and "Otherwise" = 0) of outcomes.
3. Step 3: Contributions of the Study
   - We find the best data-split which can explains better both access to loan and loan type.
   - We find using data-set as a whole shows the lowest predictive power.

Workflow

1. Step 1: Objectives of the Study
   - Which data-split can characterize better characteristics to access loan?
2. Step 2: Methodology 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 three multi logit and three multi-nominal logit models. Each data-sets were spited into 80:20 train:test groups and 10-CV were applied.
   - Logistic Regression: Access to loan.
     - We have 2 levels ("Access to Loan" = 1, and "Otherwise" = 0) of outcomes.
   - Multi Logit Regression: Access to loan type.
     - We have 4 levels ("Formal Loan" = 1, "Informal Loan" = 2, "Both Loans" = 3, and "Otherwise" = 0) of outcomes.
3. Step 3: Contributions of the Study
   - We find the best data-split which can explains better both access to loan and loan type.
   - We find using data-set as a whole shows the lowest predictive power.

Logistic Regressions

- Logistic
  - n training samples with an input vector of length d, where the d components of the input vector $\boldsymbol{x} = [x_1, ..., x_d]^T \in \mathbb{R}^d$.
  - 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

$$log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \boldsymbol{x_i}\beta = \beta_0 + \sum_{i=1}^{p} \beta_i \boldsymbol{x_i} \quad (1)$$

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 $\boldsymbol{x}$.

Logistic Regressions

- Logistic
  - n training samples with an input vector of length d, where the d components of the input vector $\boldsymbol{x} = [x_1, ..., x_d]^T \in \mathbb{R}^d$.
  - 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

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

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 $\boldsymbol{x}$.

Multi Logit Regressions

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

$$\pi_i^h = P(y_i^h = 1|\boldsymbol{x_i^h}) \tag{2}$$

  - 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

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

Multi Logit Regressions

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

$$\pi_i^h = P(y_i^h = 1 | \boldsymbol{x}_i^h) \tag{2}$$

  - 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

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

Results

	GLM.1	GLM.2	GLM.3	MLR.1	MLR.2	MLR.3
Urb & Rrl	0.673	0.679	0.667	0.685	0.684	0.678
Educ.0 & Educ.1	0.651	0.655	0.646	0.678	0.678	0.671
CCP.0 & CCP.1	0.654	0.658	0.652	0.663	0.663	0.658
Sex.0 & Sex.1	0.652	0.656	0.648	0.668	0.670	0.667
BchMk	0.655	0.658	0.653	0.660	0.658	0.657

Table: AUC's of Logit and Multi Logit models.

- Splitting the CHFS survey according to the area (Urb & Rrl) has higher predictive power and performs better to explains the Chinese household characteristics.
- Our findings line with Chen and Jin (2017); Fung´a˘cov´a et al. (2014), and indicate credit ownership is low. Formal financial inclusion is particularly constrained which is distributed economically advantaged groups (e.g., state-owned companies, affiliation with the current Chinese communist party).
- Our findings suggest that financial inclusion remains a major issue and to reduce income inequity policies must target economically disadvantage households.

Results

	GLM.1	GLM.2	GLM.3	MLR.1	MLR.2	MLR.3
Urb & Rrl	0.673	0.679	0.667	0.685	0.684	0.678
Educ.0 & Educ.1	0.651	0.655	0.646	0.678	0.678	0.671
CCP.0 & CCP.1	0.654	0.658	0.652	0.663	0.663	0.658
Sex.0 & Sex.1	0.652	0.656	0.648	0.668	0.670	0.667
BchMk	0.655	0.658	0.653	0.660	0.658	0.657

Table: AUC's of Logit and Multi Logit models.

- Splitting the CHFS survey according to the area (Urb & Rrl) has higher predictive power and performs better to explains the Chinese household characteristics.
- Our findings line with Chen and Jin (2017); Fung´a˘cov´a et al. (2014), and indicate credit ownership is low. Formal financial inclusion is particularly constrained which is distributed economically advantaged groups (e.g., state-owned companies, affiliation with the current Chinese communist party).
- Our findings suggest that financial inclusion remains a major issue and to reduce income inequity policies must target economically disadvantage households.

Results

	GLM.1	GLM.2	GLM.3	MLR.1	MLR.2	MLR.3
Urb & Rrl	0.673	0.679	0.667	0.685	0.684	0.678
Educ.0 & Educ.1	0.651	0.655	0.646	0.678	0.678	0.671
CCP.0 & CCP.1	0.654	0.658	0.652	0.663	0.663	0.658
Sex.0 & Sex.1	0.652	0.656	0.648	0.668	0.670	0.667
BchMk	0.655	0.658	0.653	0.660	0.658	0.657

Table: AUC's of Logit and Multi Logit models.

- Splitting the CHFS survey according to the area (Urb & Rrl) has higher predictive power and performs better to explains the Chinese household characteristics.
- Our findings line with Chen and Jin (2017); Fung´a˘cov´a et al. (2014), and indicate credit ownership is low. Formal financial inclusion is particularly constrained which is distributed economically advantaged groups (e.g., state-owned companies, affiliation with the current Chinese communist party).
- Our findings suggest that financial inclusion remains a major issue and to reduce income inequity policies must target economically disadvantage households.

Results

	GLM.1	GLM.2	GLM.3	MLR.1	MLR.2	MLR.3
Urb & Rrl	0.673	0.679	0.667	0.685	0.684	0.678
Educ.0 & Educ.1	0.651	0.655	0.646	0.678	0.678	0.671
CCP.0 & CCP.1	0.654	0.658	0.652	0.663	0.663	0.658
Sex.0 & Sex.1	0.652	0.656	0.648	0.668	0.670	0.667
BchMk	0.655	0.658	0.653	0.660	0.658	0.657

Table: AUC's of Logit and Multi Logit models.

- Splitting the CHFS survey according to the area (Urb & Rrl) has higher predictive power and performs better to explains the Chinese household characteristics.
- Our findings line with Chen and Jin (2017); Fung´a˘cov´a et al. (2014), and indicate credit ownership is low. Formal financial inclusion is particularly constrained which is distributed economically advantaged groups (e.g., state-owned companies, affiliation with the current Chinese communist party).
- Our findings suggest that financial inclusion remains a major issue and to reduce income inequity policies must target economically disadvantage households.

Outline

Motivation

- access to credit in rural areas are less
- urban credit has increased
- where are the poor people

Motivation

- access to credit in rural areas are less
- urban credit has increased
- where are the poor people

Motivation

- access to credit in rural areas are less
- urban credit has increased
- where are the poor people

Workflow

1. Step 1: Objectives of the Study
   - Where are the poors?
2. Step 2: Methodology Clustering households based on their financial dkajsd.
   - Unsupervised: K-means Clustering
     - We clustered households based on their financial strength.
   - Supervised: Bagging, Random Forest, and Gradient Boosting
     - After K-means clustering, we modeled the clusters to understand the accuracy of the clustering.
3. Step 3: Contributions of the Study
   - We clustered the households based on the blabla
   - We map advandaged & disadvantaged groups.

Workflow

1. Step 1: Objectives of the Study
   - Where are the poors?
2. Step 2: Methodology Clustering households based on their financial dkajsd.
   - Unsupervised: K-means Clustering
     - We clustered households based on their financial strength.
   - Supervised: Bagging, Random Forest, and Gradient Boosting
     - After K-means clustering, we modeled the clusters to understand the accuracy of the clustering.
3. Step 3: Contributions of the Study
   - We clustered the households based on the blabla
   - We map advandaged & disadvantaged groups.

Workflow

1. Step 1: Objectives of the Study
   - Where are the poors?
2. Step 2: Methodology Clustering households based on their financial dkajsd.
   - Unsupervised: K-means Clustering
     - We clustered households based on their financial strength.
   - Supervised: Bagging, Random Forest, and Gradient Boosting
     - After K-means clustering, we modeled the clusters to understand the accuracy of the clustering.
3. Step 3: Contributions of the Study
   - We clustered the households based on the blabla
   - We map advandaged & disadvantaged groups.

Unsupervised Learning

- Clustering
  - Let C_k be the generic cluster of observations with $k = 1, .., K$ which satisfies the two properties:
    - $C_1 \cup C_2 \cup ... \cup C_K = \{1, ..., n\}$, where each observation belongs to at least one cluster.
    - $C_k \cap C_{k'} = 0$ for all $k \neq k'$. Each observation belongs only one cluster.
  - K-means cluster sets the $within - cluster\ variance$ as small as possible, and it is defined as

$$\min_{C_1,...,C_K} \left\{ \sum_{k=1}^{K} W(C_k) \right\} = \min_{C_1,...,C_K} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\} \quad (3)$$

    - where $|C_k|$ is the number of observation in the k-th cluster and $W(C_k)$ is a measure of distance, called *squared Euclidean distance.*

Unsupervised Learning

- Clustering
  - Let C_k be the generic cluster of observations with $k = 1, .., K$ which satisfies the two properties:
    - $C_1 \cup C_2 \cup ... \cup C_K = \{1, ..., n\}$, where each observation belongs to at least one cluster.
    - $C_k \cap C_{k'} = 0$ for all $k \neq k'$. Each observation belongs only one cluster.
  - K-means cluster sets the $within - cluster\ variance$ as small as possible, and it is defined as

$$\min_{C_1,...,C_K} \left\{ \sum_{k=1}^{K} W(C_k) \right\} = \min_{C_1,...,C_K} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\} \tag{3}$$

    - where $|C_k|$ is the number of observation in the k-th cluster and $W(C_k)$ is a measure of distance, called *squared Euclidean distance*.

Supervise Learning

- Baging
  - FORMULA
- Boosting
  - FORMULA
- Random Forest
  - FORMULA

Supervise Learning

- Baging
  - FORMULA
- Boosting
  - FORMULA
- Random Forest
  - FORMULA

Supervise Learning

- Baging
  - FORMULA
- Boosting
  - FORMULA
- Random Forest
  - FORMULA

Results

- put tables
- clustering blabla
-

Results

- put tables
- clustering blabla
-

Results

- put tables
- clustering blabla
-

Micro App

- Micro Shinyapp[2] provides an interactive user interface, using the 2015 CHFS data-set.
- Micro Shinyappis built to map the Clusters 1 and Cluster 2&3 where Cluster 1 and Cluster 2&3 can be interpreted as low and medium & high financial healthy households.
- Policy makers can see the distribution of the survey population based on the selected criteria (e.g., Clusters 1 and/or Cluster 2&3) to design suitable programs, aiming the most poor ares primarily.
- Map of the households can be found under the Map tab and Table tab prints out the pivot table.

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

Micro App

- Micro Shinyapp[2] provides an interactive user interface, using the 2015 CHFS data-set.
- Micro Shinyappis built to map the Clusters 1 and Cluster 2&3 where Cluster 1 and Cluster 2&3 can be interpreted as low and medium & high financial healthy households.
- Policy makers can see the distribution of the survey population based on the selected criteria (e.g., Clusters 1 and/or Cluster 2&3) to design suitable programs, aiming the most poor ares primarily.
- Map of the households can be found under the Map tab and Table tab prints out the pivot table.

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

Micro App

- Micro Shinyapp[2] provides an interactive user interface, using the 2015 CHFS data-set.
- Micro Shinyappis built to map the Clusters 1 and Cluster 2&3 where Cluster 1 and Cluster 2&3 can be interpreted as low and medium & high financial healthy households.
- Policy makers can see the distribution of the survey population based on the selected criteria (e.g., Clusters 1 and/or Cluster 2&3) to design suitable programs, aiming the most poor ares primarily.
- Map of the households can be found under the Map tab and Table tab prints out the pivot table.

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

Micro App

- Micro Shinyapp[2] provides an interactive user interface, using the 2015 CHFS data-set.
- Micro Shinyappis built to map the Clusters 1 and Cluster 2&3 where Cluster 1 and Cluster 2&3 can be interpreted as low and medium & high financial healthy households.
- Policy makers can see the distribution of the survey population based on the selected criteria (e.g., Clusters 1 and/or Cluster 2&3) to design suitable programs, aiming the most poor ares primarily.
- Map of the households can be found under the Map tab and Table tab prints out the pivot table.

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

Thank you