DETERMINANTS OF FINANCIAL INCLUSION - THE VIETNAMESE HOUSEHOLDS' PERSPECTIVE

A Research Paper presented by:

Ha Nguyen Son
(Vietnam)

in partial fulfillment of the requirements for obtaining the degree of
MASTER OF ARTS IN DEVELOPMENT STUDIES

Major:

ECONOMICS OF DEVELOPMENT
(ECD)

Specialization:

Econometric Analysis of Development Policies

Members of the Examining Committee:

Professor Arjun Singh Bedi
Professor Vo Hong Duc

The Hague, The Netherlands
December 2022

Disclaimer:

This document represents part of the author's study programme while at the International Institute of Social Studies. The views stated therein are those of the author and not necessarily those of the Institute.

Inquiries:

International Institute of Social Studies
P.O. Box 29776
2502 LT The Hague
The Netherlands

t: +31 70 426 0460
e: info@iss.nl
w: www.iss.nl
fb: http://www.facebook.com/iss.nl
twitter: @issnl

Location:

Kortenaerkade 12
2518 AX The Hague
The Netherlands

Acknowledgments

I would like to express my warm gratitude to Professor Arjun Singh Bedi and Professor Vo Hong Duc for their kind guidance, comments, and feedback during the progress of finishing my research paper. Your detailed and conscientious guidance, especially in the methodology part, truly helped me a lot to attain consistent and reportable empirical results. Without your empathy guidance, the research paper will undoubtedly not have been throughout completed.

During the unforgettable journey, I would also like to say warm thanks to Professor Matthias Rieger and respectable ISS professors and staffs for their comprehensive support and assistance in my two semesters studying at ISS. The living experiences and lessons that I have learned from ISS in the Netherlands are undoubtedly one of the best things in my life that I have ever experienced.

I would also like to thank my friends at ISS, especially three of my close friends: Tom (Long Hoang), Charlie, and Anvisha, for their lovely support and assistance during the time I lived in the Netherlands. The good memories with ISS's beloved friends will always be in my heart.

Finally, words can not express my gratitude for my small family. Thank you, Mom and beloved sisters and brothers, for your unconditional love and support during the time I lived in the Netherlands, which also was the first time in my life that I went to a place outside of Vietnam.

Contents

List of Tables

List of Figures

List of Maps

List of Appendices

List of Acronyms

ISS	Institute of Social Studies
UNDP	United Nations Development Programme
IMF	International Monetary Fund
IFI	Index of financial inclusion
LFI	Level of financial inclusion
OLS	Ordinary Least Squares
OLM	Ordinal (Ordered) Logit Model

Abstract

Using the Vietnam Household Living Standard Survey (VHLSS) datasets in 2014, 2016, and 2018, this study examines financial inclusion's key determinants from the perspective of Vietnamese households. This study investigates the effects of the determinants on each level of financial inclusion utilizing an ordered logit model. The empirical results reveal that total income per household, relative income index, and the distance of the household from the nearest financial hub are crucial factors that drive the financial inclusion level. While the total income per household plays a positive role in enhancing financial inclusion, relative income seems to have an adverse effect on the degree of financial inclusion of individuals. Besides, distance to the nearest public bank branch is also a challenge to overcome in achieving the financial inclusion goals in Vietnam in the future.

Relevance to Development Studies

The World Bank (2012) considers that financial inclusion stands as one of the crucial mechanisms to promote economic growth, reduce poverty, and overcome income inequality. By investigating the determinants of financial inclusion using micro-data in the Vietnamese context, this research contributes to the existing literature by providing a new perspective on Vietnamese households' characteristics. The findings of this research can also be helpful for scholars and policymakers regarding the new angles on the main factors that drive financial inclusion in the developing country. Policy implications have emerged based on the findings to attain the goals of financial inclusion.

Keywords

Financial inclusion, Income inequality, Relative income, Household characteristics, Vietnam.

Chapter 1
Introduction

1.1. Contextual background

Over the past decades, the literature has highlighted the multiple roles of financial inclusion, such as decreasing poverty, achieving inclusive growth, maintaining the inclusion of authorization policies, and strengthening the financial sector's sustainability. The access of low-income households' small and middle-sized companies to financial services would considerably increase their propensity to engage in future purchases, as would the availability of banking services (Sarma & Pais, 2011; Kim et al., 2018). Furthermore, Li (2018) explains that the ease of financial market involvement has various positive effects on the well-being of disadvantaged family members. Moreover, by promoting digitalized payments and transactions, financial inclusion can efficiently improve the allocation of financial resources for insufficient government programs and reduce the cost of waiting, traveling, and other expenses. It is also acknowledged that financial inclusion can boost banking institutions with low-cost and steady deposits, which constitute a solid foundation for lending operations. Additionally, low-income savers and borrowers can sustain their spending habits by maintaining deposits and taking loans.

In addition, financial inclusion reduces poverty through promoting financial system activities, including investments and savings. Investments and savings, in general, are vital instruments that allow all sectors of society to engage in the official financial system. Increasing the number of persons participating in unofficial economic systems also contributes to developing more efficient monetary policies. Therefore, enhanced financial inclusion, a solid strategy, and the right political direction can reduce poverty and promote economic growth. Regarding the advantages of financial inclusion for enterprises and individuals, governments worldwide have established various programs to promote financial inclusion.

Financial inclusion has been widely regarded as an essential tool for promoting economic growth and reducing poverty in Vietnam over the past decade. After more than three decades since the "Doi Moi" economic reform in 1986, Vietnam has transitioned from one of the poorest countries to a lower middle-income country, thanks to the Vietnamese government's various economic policies (see Figure 1.1 and Appendix 1). However, along with economic growth, the sustainability of economic growth and the distribution of income across Vietnam are among the most controversial topics that need to be addressed. However, there are still limited studies on the crucial factors that enhance financial inclusion, especially at the household level in Vietnam. To contribute to the existing literature, I am interested in investigating the key determinants of financial inclusion from the perspective of Vietnamese households, mainly focusing on the income level, the relative income at the provincial level, and the distance to the closest financial hub. Precisely, the main focus of this research will be to investigate the role of financial inclusion in reducing income inequality in Vietnam.

1.2. Research objectives and research questions

Many countries have attempted to improve financial inclusion levels throughout the past decade by creating banking systems and services accessible to all economic participants. Literature has highlighted the various roles of financial inclusion, such as improving people's lives, promoting inclusive economic growth, ensuring the effectiveness of welfare programs, and providing financial sector sustainability. However, as discussed in the previous chapter,

limited studies try to investigate the key factors driving financial inclusion in Vietnam. This research work was motivated by ongoing practical controversy and the lack of empirical research on this subject.

Based on the problem statement and the literature, the main research question and three sub-research questions will be examined as elaborated below:

The main research question:

- What are the key determinants driving financial inclusion in Vietnam?

Three sub-research questions:

- What is the link between financial inclusion and income level?
- What is the connection between financial inclusion and relative income?
- Does distance to the nearest financial hub/bank branch affect houseold financial inclusion in Vietnam?

For each sub-research question, three new variables: The index of financial inclusion (IFI), the relation income (RI), and the distance from the household's nearest bank branch (DNB), will be constructed based on the existing literature.

1.3. Limitations of the study

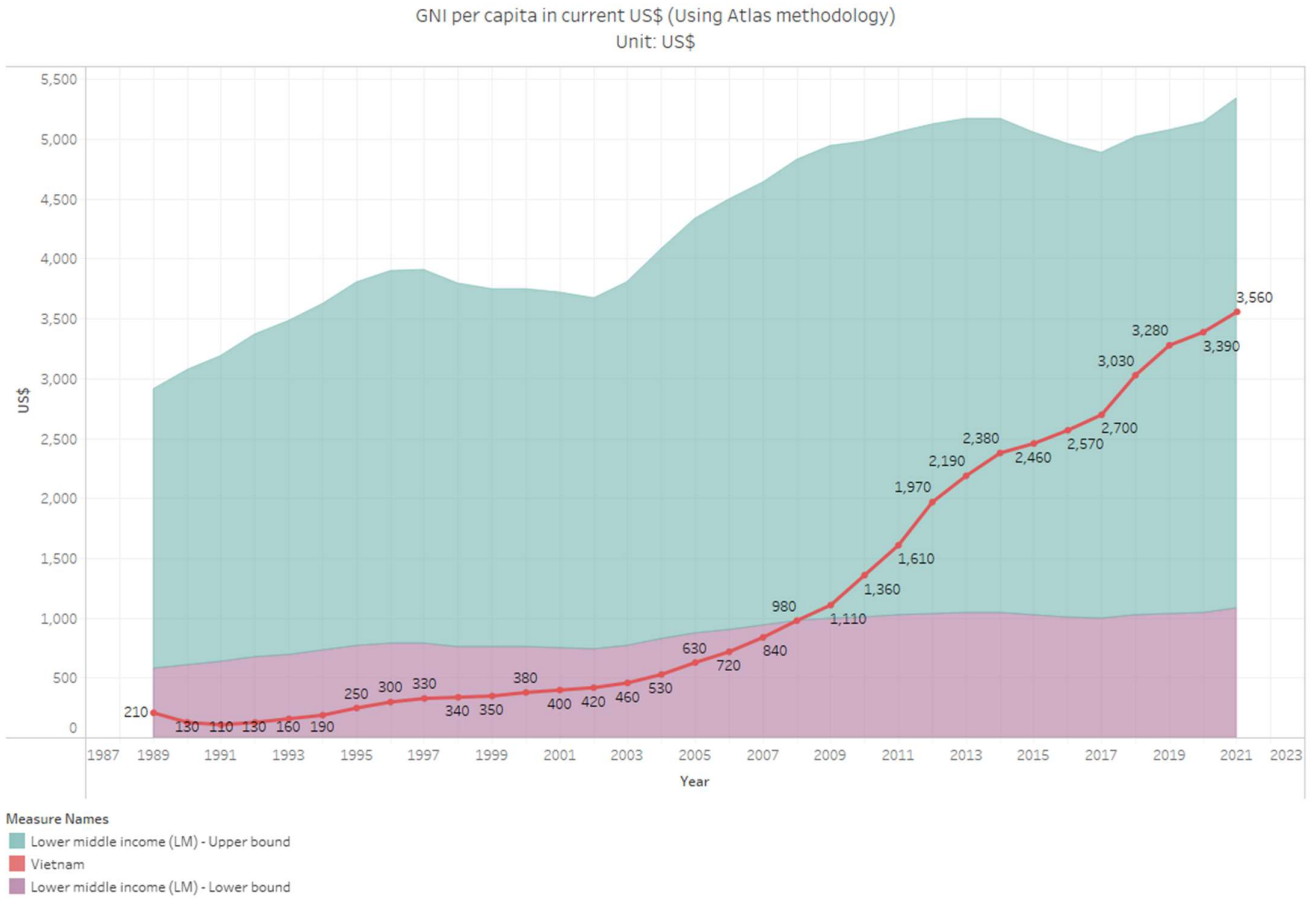
Although in this paper, relatively large sample sizes are employed in the analysis/regressions (34,437 observations in 2018, 43,459 observations in 2016, and 7,979 observations in 2014), we can only observe the effects of the relationship based on cross-sectional data across years. Due to the nature of VHLSS data, only 50% of surveyed households will be invited for the following VHLSS survey (which is conducted every two years). This creates a significant problem because the sample size will be reduced to at least 50% every time we merge the datasets. Moreover, since the interpretation and post-test estimation of fixed and random effects in the ordinal logit regression is still a controversial topic for debate, I decided not to merge these datasets into panel data in order to run a panel regression. Three cross-sectional regressions will be provided separately. Therefore, this paper might not observe the time-effect and causality relationships between variables. Future studies can utilize the results from this paper to empirically examine the time-varying effects of critical determinants of financial inclusion using other panel regression methodologies.

1.4. The structure of the research paper

The paper is structured as follows. Following the Introduction, Chapter 2 reviews key literature about the determinants of financial inclusion, both at the global level and in Vietnam, as well as the current trend of empirical findings about the relationships among financial inclusion, income, and income inequality. Chapter 3 provides the theoretical framework to theoretically explain the reasons behind selecting each determinant in the main equation based on the mentioned literature review. Chapter 4 shows the methodologies and descriptions of datasets used in the regression. Chapter 5 provides pieces of empirical evidence based on OLS and OLM regression results. Finally, Chapter 6 concludes by reviewing the main findings and suggests several policy implications to attain the financial inclusion goal.

Figure 1.1

The transition of Vietnam to a lower middle-income country



Source: World Development Indicators database, visualizating by the author using Tableau (Accessed: November 5th 2022)

Chapter 2
Literature review

2.1. Definitions of Financial Inclusion and its measurements

2.1.1 Financial Inclusion and its measurements in the world

Over the last few years, various countries have been trying to achieve financial inclusion by developing financial systems and services to approach every user in the economy. Literature has identified the multiple roles of financial inclusion, such as reducing poverty, achieving inclusive economic growth, assuring government policies' extensiveness, or offering elasticity to the financial sector. According to Sahay et al. (2015) from IMF, over 60 nations have set financial inclusion as one of the main goals to support their economic growth and development.

In general, definitions of Financial Inclusion and some methods of calculating financial inclusion indices have been widely introduced in previous studies. According to Sarma (2008, 2012), Financial Inclusion is a mechanism that guarantees all society participants may easily access, use, and benefit from the official financial system. In this paper, Sarma introduces a unique measurement called the Index of Financial Inclusion (IFI). IFI is a multidimensional indicator with a similar calculating approach compared to other famous economic development indexes, such as the Human Development Index (HDI) or Gender Development Index (GDI). IFI combines various banking sector indicators and captures them into one number ranging from 0 to 1, with 0 denoting complete financial exclusion and 1 denoting complete financial inclusion in a particular country. Many scholars use this approach to measure IFI since it indicates three main dimensions: Ease of access, Availability, and Usage, which can be easily computed using simple weighted average value methods. The selection of these three main dimensions of Sarma is based on another famous study by Demirguc-Kunt & Klapper (2012): The Global Findex Database, which will be briefly reviewed later. A notable difference between the proposed IFI versus other mentioned indexes (HDI, GDI) is that Sarma tries to calculate each dimension as well as the completed index based on the concept of Euclidean distance between the point of completed financial exclusion (i.e., the point that takes the value of 0) and the point of completed financial inclusion (i.e., the point that takes the value of 1). Because of the simple mathematical approach of this proposed method, this index can also be used to measure the financial inclusion level of micro-level data like provincial data or village data. After calculating the IDI for 94 countries from 2004 - 2010, Sarma (2012) also found that the financial inclusion level and the average income level tend to move in the same direction in most countries.

As mentioned above, one of the most famous notations of financial inclusion comes from the study of Demirguc-Kunt & Klapper (2012), since they tried to construct a set of official indicators to measure financial inclusion across nations, namely the Global Findex Database. This database is collected from over 148 countries worldwide, including low-income, middle-income, and high-income countries. The Global Findex Database focuses on financial services usage, not on the access to financial services (Commonly known as financial services penetration). Unlike financial services penetration, which is heavily affected by the supply-side (i.e., financial service providers), financial services usage is affected by both the supply-side and demand-side (i.e., financial service users). In this database, they divide their collected indicators into three sub-groups: Indicators of account penetration, indicators of saving, and indicators of borrowing. This study contributes to the literature by filling various gaps, such

as the lack of systematic financial inclusion data across countries and the incomplete split of the aggregated financial inclusion data across categories. Up to the present time, this project is still updated and provided every three years, with the last version being published in 2021. Partly motivated by this study, many scholars try to propose new methods to capture the multidimensional characteristic of financial inclusion. Noelia & Tuesta (2014) propose new methods to measure the level of financial inclusion mathematically. In their paper, they point out two critical weaknesses in the current method of calculating the multidimensional financial index. Firstly, existing financial inclusion index calculations are heavily based on supply-side data (commonly measured by the number of loans or accounts of users). However, these measures sometimes are inaccurate since one person can easily have more than a bank account or loan amount. Secondly, the assignment of weights for each dimensional indicator of financial inclusion: Access, Barriers, and Usage is mainly based on subjective pieces of evidence and a lack of mathematical concepts. They propose another unique way to calculate the financial inclusion index to build a better index that is responsive to all of the above problems. Their improved financial index uses the two-step PCA (i.e., Principal Component Analysis) to calculate the weight of each sub-dimensional index's weight and compute the final index based on the weighted average method of Sarma (2008, 2012). After employing this approach to calculate the financial inclusion index for 82 countries in 2011, they found that "Access" gets the highest weight among the three sub-indexes. In other words, key factors can determine the level of financial inclusion, which means that the official financial services supply (for instance: Commercial banks) plays a more critical role in determining the financial inclusion level than the number of users.

After Sarma introduces the financial inclusion index, growing literature attempts to improve the techniques of allocating the weights of each sub-indexes. Amidzic et al. (2014) also notice shortcomings of the existing approach to measuring financial inclusion. They aim to propose their index in order to overcome all of the mentioned shortcomings mathematically. This paper introduces two dimensions of financial inclusion versus three in Sarma's studies. These two dimensions are the capacity of financial services and the financial services usage. This approach is mainly different from other indexes since it uses the Factor Analysis approach to group sub-dimensional indexes into the most appropriate dimension. Besides, using this approach, the randomly assigned weights based on subjective choices of other studies are well-replaced by a more mathematical way of weights assignment. Similarly, to prevent subjective weighting issues, Wang & Guan (2016) employ the Coefficient of Variation method to objectively distribute the weights among sub-dimensional indexes. After calculating the indexes for 127 countries worldwide in 2011, the geographical distribution of financial inclusion has been found: Most Asian and African countries have lower financial inclusion levels than European and North-American countries. Moreover, their spatial econometric regression shows those essential elements that indicate a person's level of financial inclusion are their income level, educational background, and technology usage. In contrast, financial depth and banking health status are the key drivers leading to increased financial inclusion.

From another perspective, Svirydzenka (2016) introduces a broad-based method, aggregating all sub-indices into the set of financial development indicators, using familiar PCA and weighted average approaches. Financial development, a more general definition, holds many similarities in terms of definitions compared to financial inclusion (Demirguc-Kunt & Klapper, 2012). Some notable indexes that can be mentioned here are indexes of financial development, financial markets, and financial institutions. While still having many limitations regarding data availability and mathematical approaches, this study sheds essential insights into various sub-factors that can affect the overall financial development level. Since the dimensions of financial inclusion are often unobservable, there is no direct way to measure this index quantitatively (Noelia, 2017). An updated version of Noelia & Tuesta's (2014) 's

paper attempts to calculate the financial inclusion index in 2014 using a two-step PCA. In terms of time effects, the financial inclusion growth from 2011 to 2014 is also provided in this paper. Most countries in the list of 137 countries have seen an increase in financial inclusion levels over time.

2.1.2 Financial Inclusion and its measurements in Vietnam

In Vietnam, the literature on financial inclusion and its measurements are growing over time. However, studies about measuring financial inclusion using micro-level data are still limited. The main barrier is data availability. We do not have enough micro-level data to measure all the dimensions of financial inclusion. Nguyen et al. (2021) are among the first Vietnamese authors to create a unique IFI using Vietnamese households' data in their study. They employ a set of Yes/No questions about the financial status of households from section 8 of the Vietnam Household Living Standards Surveys 2018 (VHLSS 2018) to calculate a unique index by province. In this paper, they find empirical evidence that the financial inclusion level of Vietnam largely depends on income distribution at the provincial level. Moreover, the crucial economic cities, especially in the North and South regions of Vietnam, tend to have higher financial inclusions than other areas. Another study from Tuyen & Van (2021) tries to measure financial inclusion with household-level data, following the guidelines of the World Bank (2018). Instead of aggregating all indicators into one single index, four-dimensional indicators of financial inclusion: Loans from formal institutions, the value of this loan, having an official bank account, and having an official saving account are employed in their model individually. They provide evidence that households living in provinces with better institutional quality tend to receive official loans, open new debit or saving accounts, and have greater access to financial opportunities.

As mentioned above, limited literature tries to construct the financial inclusion index using household-level data in Vietnam. The emergence of an ongoing practical debate about measuring financial inclusion using micro-level data and a scarcity of empirical studies on the significant linkage between financial inclusion and income inequality in Vietnam are the main inspirations for us to start this research paper.

2.2. The relationship between Financial Inclusion, income and income inequality

With the development of Financial Inclusion measurement literature, there is also a vast literature about income inequality and its relationship with financial inclusion that has continued to grow over the past decades. As one of the most notable studies that attempt to investigate the effect of income distribution on the stability of macroeconomics, Galor & Zeira (1993) demonstrate that given existing market imperfections, the distribution of income can heavily affect investment, spending, and saving behaviors individually, thereby affecting the aggregate activity of the economy. This study stands as one of the foundational papers that motivates other scholars to examine this relationship further. Demirguc-Kunt & Levine (2009) propose a systematic review of the relationship between finance in general and income inequality to summarize the growing body of literature. According to their extensive review, in most studies they examined, financial development plays a crucial role in lowering the level of inequality, both at national and household levels. Besides, the authors highlight a significant gap in existing literature up to that moment, that they failed to empirically confirm the mutual relationship between growth, inequality, and financial development. Another country-level study finds that financial development negatively correlates with inequality, with financial literacy level acting as an underlying driving factor (Prete, 2013). This result is

also consistent with the research of Balakrishnan et al. (2013) since they also confirm that the development of the financial system and the increasing degree of financial inclusiveness contribute to overall economic growth and income distribution in Asia countries. Notably, Balakrishnan et al. (2013) point out another exciting angle from their study: the degree of this contribution of financial inclusiveness to the even income distribution is not fixed but greatly varies across nations and regions. Similarly, using a dataset of 37 Asian countries, Park & Merkado (2015) find empirical evidence that financial inclusion significantly reduces income inequality and poverty in general. However, the level of this relationship heavily depends on the demographics of countries, such as population. Another research from Kim (2016) emphasizes the positive and crucial role of financial accessibility, represented by financial inclusion, in lowering the inequality level, especially in low-income nations. Furthermore, together with moderating the income inequality level, they surprisingly examine that financial inclusion can eventually change the common negative sign of the relationship between growth and income inequality to a positive sign in the long run. However, not every study confirms the negative relationship between financial development and income inequality. Using a panel fixed-effects analysis of 138 countries in 48 years, Jauch & Watzka (2011) interestingly suggests that income inequality becomes more severe with the development of finance.

Up to this point, we have run through some notable studies using country-level data to examine the mentioned relationships. From another research angle, using micro-data level, various studies also confirm the role of financial inclusion in improving income levels and tackling unequal income distribution among individuals. Using micro-level data from the National Employment survey to examine the case study of Banco Azteca, a well-known bank in Mexico, research from Miriam & Inessa (2014) suggests that the increasing number of opened bank branches, which is one of the vital dimensions of financial inclusion (Sarma, 2008, 2012; Demirguc-Kunt & Klapper, 2012) can positively affect the income level of individuals and tackle poverty issues. Similarly, utilizing the national finance survey data of over 6,000 households in China, Zhang & Alberto (2017) highlight the importance of enhancing financial inclusion in improving people's living standards, reducing poverty, and reducing income inequality in the context of Chinese households. Another interesting thing in this study is that the authors find that the group of low-level income households benefits more from enhancing financial inclusion progress than the wealthier ones. This empirical evidence suggests that enhancing financial inclusion can ameliorate income inequality issues.

With the growing body of literature about financial inclusion and its relationships with income and income inequality, this topic also draws considerable attention from researchers and policymakers in Vietnam. However, limited literature aims to investigate this relationship directly due to data availability and the difficulties when constructing the micro-level financial inclusion index. A common approach is using one dimension of financial inclusion and then investigating its relationship with income and income inequality. To begin with, the research of Mikkel & Finn (2008) is among the first studies that try to investigate the attributes of the credit market and its demand (both in formal and informal sectors), which is one of the well-known determinants of financial inclusion (Sarma, 2008, 2012; Demirguc-Kunt & Klapper, 2012; Noelia & Tuesta, 2014; Amidzic et al., 2014). Using data from Access To Resources survey in 2003 and VHLSS in 2023, including more than 900 households across four provinces in Vietnam, they find empirical evidence that access to credit, as well as the credit demand, depends heavily on geographical and regional aspects. While people with higher education backgrounds and living in urban areas have more opportunities to access formal credit, the informal credit market seems to attract more people from rural areas and lower education backgrounds. Another paper by Luan & Bauer (2016) confirms the hypothesis that access to credit levels varies differently between rural and urban areas. Besides, they find that increasing credit access can significantly boost total household income in rural areas,

particularly in the Northern area of Vietnam. From a different perspective, Nguyen et al. (2018) investigate the key determinants of preferential loans from the Vietnamese households' perspective, using the Vietnam Household Living Standards Survey (VHLSS) in 2010, as well as its effect on the total income of each household. In general, access to preferential loans and access to loans is also well recognized as one of the main dimensions of financial inclusion (Noelia & Tuesta, 2014). Unexpectedly, Nguyen et al. (2018) suggest that access to preferential loans is not a factor that contributes to the increase in households' income.

To better capture the critical information of the literature which has been mentioned above, a table of summarized literature with coefficient signs of the relationship between financial inclusion, income, and income inequality will be illustrated in Table 2.1 below:

Table 2.1

A summary of relevant literature

Author	Data & Timespan	Method	Sign of relationship between financial inclusion and:		Other relevant findings
			Income	Income inequality	
Macro-level data					
Sarma (2010)	47 countries (2004)	OLS	Positive	Negative	There are positive relationships between financial inclusion and urbanization. Literacy level plays a significant role in improving financial inclusion.
Jauch & Watzka (2011)	138 countries (1960 – 2008)	OLS, 2SLS	Positive	Positive	The effect of financial inclusion on decreasing income inequality levels is not significant in low-income countries.
Prete (2013)	30 countries (1980 – 2005)	2SLS		Negative	Financial development and the growth of income inequality are negatively correlated.
Park & Mercado (2015)	37 developing Asian countries (2004 – 2012)	OLS, with average value of variables from 2004 to 2012	Positive	Negative	The effect of educational level on financial inclusion is not significant.
Kim (2016)	40 OECD and EU countries	Panel fixed-effects, GMM		Indirectly negative (In low-income countries)	Financial inclusion can act as an indirect factor that can accelerate the relationship between income inequality and economic growth and change this relationship from negative to positive.
Wang & Guan (2016)	75 countries	OLS, Spatial regression	Positive	Negative (But not significant)	Unemployment ratio, Literacy level, and gender (Being male) are essential factors that enhance financial inclusion.

Micro-level data					
Zhang & Alberto (2017)	6,195 Chinese households (2011)	Quantile regression, OLS	Positive	Negative	The household head's gender (Male), married status, and educational level positively boost financial inclusion. Low-income households seem to benefit more from financial inclusion than richer ones.
Mikkel & Finn (2008)	932 households from VHLSS 2022 and VARHS 2022	OLS, Probit model			People with higher education backgrounds and living in urban areas have more opportunities to access formal credit. The informal credit market attracts more people from rural areas and lower education backgrounds. While the distance to the village center is negatively correlated with credit access, the household head's age and gender effects on credit demand are mixed.
Luan & Bauer (2016)	1,338 households in VARHS 2012	Probit model, propensity score matching	Positive		Increasing credit access can significantly boost the total income of households in rural areas. Younger household heads tend to get credit access more accessible than older ones.
Nguyen et al. (2018)		Quantile regression			Access to referential loans is not a critical factor contributing to the increase in household income.

Source: Summarized by the author

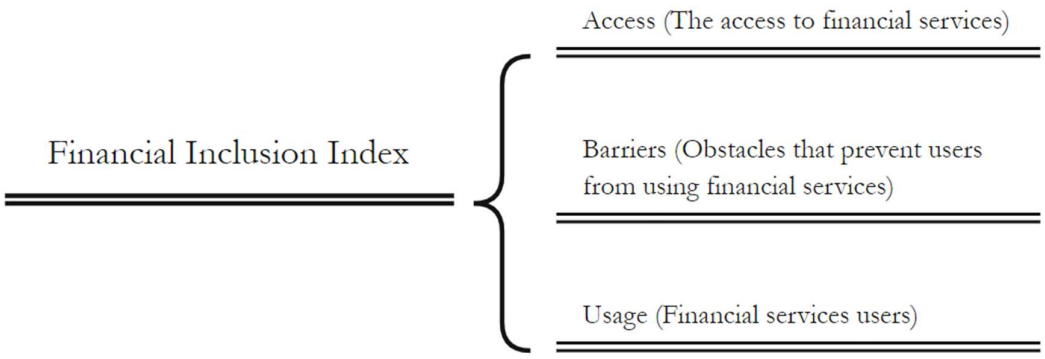
Chapter 3
Theoretical framework

3.1 An index of Financial Inclusion

Financial inclusion has been remarked as a multidimensional definition (Sarma, 2008, 2012; Demirguc-Kunt & Klapper, 2012; Noelia & Tuesta, 2014, 2017; Amidzic et al., 2014; Svirydzenka, 2016). As mentioned above, after Sarma's introduction of one of the first financial inclusion indexes, various literature has focused on improving the calculation methods of this index. Among these studies, Noelia & Tuesta (2014, 2017) and Svirydzenka (2016) propose the same conceptual framework for computing the degree of financial inclusion. Following past literature, in their paper, the financial inclusion degree is shaped by three-dimensional aspects: Access (i.e., access to financial services), Barriers (Obstacles that prevent users from starting using financial services), and Usage (Financial services users). (See Figure 3.1)

Figure 3.1

The well-known dimensional components of financial inclusion



Source: Noelia & Tuesta (2014, 2017) and Svirydzenka (2016)

Typically, with macro-level data (For instance: Global Findex Database), we can efficiently compute the "Access" component, which is a supply-size dimension. The "Access" dimension is commonly constructed using three sub-indicators: the number of ATMs per 100,000 adults, the number of official financial institutions per 100,000 adults, and the number of financial providers per 100,000 adults (Sarma, 2008, 2012; Noelia and Tuesta, 2014, 2017; Svirydzenka, 2016). However, considering the data availability of household-level data in Vietnam, we cannot find appropriate supply-side data to measure this dimension. Hence, to measure the financial inclusion index, we partly follow the conceptual approach in Figure 1.1. On the one hand, we construct the index based on two remaining dimensions: Barriers and Usage, which are demand-size dimensions (Svirydzenka, 2016). On the other hand, we utilize the collection of 7 questions in Section 8: Participating in aid schemes status of households in the Vietnam Household Living Standards Survey (VHLSS) dataset to calculate the "Usage" dimension, similar to the approach of Nguyen et al. (2021). However, instead of taking an arithmetic mean of the financial inclusion index and calculating it at the provincial

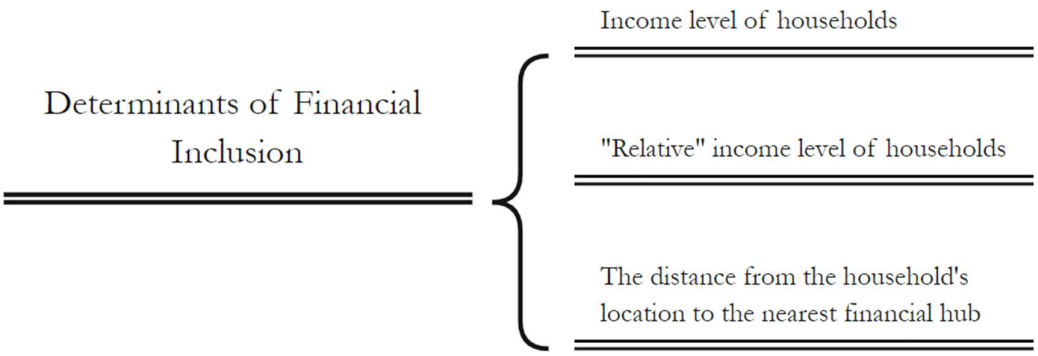
level like them, in this paper, we propose a different way to construct the financial inclusion index, with will be described in detail in Section 4.3 below.

Moving to the third dimension: Barriers. we aim to set it as one of the independent variables purposely. This variable is one of the unique approaches of this paper compared to other studies since it is the only variable we do not utilize from the VHLSS dataset. This is also why we do not directly put this dimension into the final financial inclusion index and the variable description like other dimensions.

Besides constructing the financial inclusion index, the primary purpose of this paper is to investigate some possible determinants that determine the level of financial inclusion in Vietnam. As mentioned in Section 1.2 above, three key determinants that will be employed in the econometrics model of this paper are the income level of households, the "relative income" level of households, and the distance from households' location to the nearest official banking institution. (See Figure 3.2)

Figure 3.2

The (proposed) determinants of financial inclusion



Source: Proposed by the author

The theoretical concept of each determinant will be described in Section 3.2 below.

3.2 The determinants of Financial Inclusion

3.2.1 Income level

Over the past decades, various studies have examined the relationship between financial inclusion and income at individual and country levels. In their study, using data from more than 6,000 Chinese households, Zhang & Posso (2017) find that the increase in financial inclusion level can lead to household income growth. Similarly, Sarma (2011) also finds that income level plays a vital role in adjusting financial inclusion, income inequality, access to information technology, and educational and financial literacy levels. However, since it is worth noting that the level of financial inclusion can be different among Asian countries (Fungacova & Weill, 2014), the necessity of determining whether income level is one of the determinants of financial inclusion should be examined using the individual and household levels in Vietnam.

In this research paper, we utilize the data from Section 4: Income of households in Vietnam Household Living Standards Surveys (VHLSS) in 2012, 2014, and 2018 to calculate the income level of each household, respectively. Since VHLSS datasets do not directly provide the aggregated income variable of households, we need to calculate it manually based on available sources of income in terms of profit and loss at the household level. The precise method to measure this determinant will be described in Section 4.3 below.

3.2.2 "Relative income" (Tunnel effect versus Joneses effect)

The definition of relative income (or "relarive deprivation" in some papers) was first introduced by Brady and Friedman (1947). It refers to the ratio of individual income to the average income in his or her compared group. In this paper, they also note that the income distribution and relative income distribution positively correlate with people's saving routine. Motivated by this study, numerous publications have examined the impact of relative income on other socioeconomic variables. Among these studies, Hirschman has been remarked as the introducer of the term "Tunnel effect" based on the notion of relative income. The tunnel effect relates to the tendency of one individual to believe in his or her income rise in the future if he or she witnesses the income rise from his or her social groups (For instance: An income rise from his or her friends, neighbors, or relatives). As a result, this individual will have the subjective expectation that there will be his or her turn to get an income increase in the future. On the other hand, Hirschman also mentions the definition of the "reversed" tunnel effect: If we witness all of the suffering, difficulties, and failures from our social groups, we tend to have a similar depression, as it may happen to us in the future as well.

Apart from the tunnel effect, another notion is "The Joneses effect." This definition comes from the sentence: "Keeping up with the Joneses," which is an English idiom that refers to the natural comparison between one individual and his or her neighbors in terms of income or value possessions, and this individual will try to purchase goods and services that his or her neighbors have, in order to "keep up" with them (Christen & Morgan, 2005). For instance, if some neighbors purchase a new iPhone, we will also want to purchase this brand-new one.

After the tunnel effect and the Joneses effect's introduction, various studies have employed the relative income variables in their research to investigate whether the tunnel effect or the Joneses effect holds. Li (2018), in his research about the relationship between the poverty level and financial inclusion level in China, confirms the presence of the tunnel effect but not the Joneses effect. In contrast, from the perspective of Vietnamese households, Binh et al. (2021) find that relative income (or "comparison income" in their paper) is negatively correlated with households' overall well-being, which suggests that the Joneses effect is dominant here.

As listed above, these two effects on the financial inclusion level and the households' well-being are mixed. Hence, in this research paper, we will employ the relative income variables to examine the effect of relative income on the level of financial inclusion in Vietnam, but with some minor changes compared to past studies to fit with the data of this paper. The straightforward way of measuring the relative income using VHLSS data will be comprehensively expressed in Section 4.3 below.

3.2.3 Distance to the nearest financial hub

According to Demirguc-Kunt & Klapper (2012), physical distance to the nearest financial institution is one of the main barriers that around fifty percent of adults worldwide face – that is, do not have bank accounts. Besides, they point out that distance from a bank is a more significant obstacle in rural areas than in urban areas. Moreover, in developing

countries, the problem of long distances between the location of unbanked persons and their nearest official bank is more severe compared to developed countries (Noelia, 2017). Based on the above empirical evidence, in this paper, we include a control variable representing distance of a particular household to their nearest official bank to investigate whether this is one of the critical factors that drive financial inclusion, as stated in the existing literature. This variable is different from other variables in this paper since this is the only variable that is obtained from outside the VHLSS datasets. Similar to these two mentioned determinants, the details of this variable and its way of collecting data will be precisely described in Section 4.3 below.

Chapter 4
Data and methodology

4.1 Econometric model – The Ordered Logit Model and its assumptions

To investigate determinants of financial inclusion in Vietnam from the household perspective, an econometric equation will be constructed as follows:

$$\boldsymbol{LFI_i} = \beta_0 + \beta_1 Age_i + \beta_2 Age^2{}_i + \beta_3 Gender_i + \beta_4 Marriage_i + \beta_5 Employed_i + \beta_6 Edu_i + \beta_7 HRI_i + \beta_8 Distance_{i\,t} + \beta_9 log(TI_i) + \beta_{10}\, Location_i + u_i$$

Where:

LFI_i denotes the level of financial inclusion of household i

Age denotes the age of the household head

Gender denotes the gender of the household head (0 - Male; 1 - Female)

Marriage denotes the marriage status of the household head of household i (0 - Other; 1 - Married)

Employed denotes the employment status of the household head of household i (0 - Unemployed; 1 - Employed)

Edu denotes the educational level of the household head of household i

HRI denotes the relative income of the household i

TI denotes the total income of all individuals in the household i

Distance denotes the nearest distance from household i to the nearest official bank

Location denotes the area where the household currently lives (0 – City/urban area; 1 – Rural area)

Since the level of financial inclusion (LFI) will be categorized into three levels: Low, Medium, and High (the details of these categorized methods will be mentioned in Section 4.3 below), we will have an ordered dependent variable, which takes three values: 1 – Denotes low level of financial inclusion, 2 - Denotes a medium level of financial inclusion and 3 - denotes a high level of financial inclusion. Since proportional odds logistic regression is the most common approach to deal with ordinal dependent variables among various types of the ordered logistic regression model (Williams, 2006; Frank, 2015), we will employ the approach of the (partial) proportional odds model in this paper.

First, let's assign the value for the financial inclusion level of Y as below:

$$\boldsymbol{y} = \begin{cases} 1 \;\; Low\ LFI \\ 2 \;\; Medium\ LFI \\ 3 \;\; High\ LFI \end{cases} \qquad (1)$$

Following the Ordinal Logit model explanation of Williams & Quiroz (2019), we will mathematically express the equation (1) as follows:

$$y_i = \begin{cases} 1 & if\ y_i^* \le \kappa_1 \\ 2\ if\ \kappa_1 \le y_i^* \le \kappa_2 \\ 3 & if\ y_i^* \ge \kappa_2 \end{cases} \quad (2)$$

Where:

y_i denotes the observed ordered LFI in household i

y_i^* denotes the unobserved variable (or latent variable) with determine the value of y_i

κ_1 and κ_2 denote the threshold points (or cut-off points)

The observed variable y is also the function of the latent (And unobserved) variable y^*. Considering the total population, we have the estimated equation for the latent variable as below:

$$y_i^* = Z_i + \varepsilon_i = \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \quad (3)$$

Where:

Z_i denotes the part of the ordinal logit model's equation

β denotes the parameters that need to be estimated

ε_i denotes the error term of the equation

Since the error terms ε_i of the above equation follows the logistic distribution with mean $\mu = 0$ and variance $s = \frac{\pi^2}{3}$, we can estimate its probability distribution and plug it on the equation later. Hence, we can remove it from the estimated equation. Then the equation (3) can be expressed as:

$$Z_i = \sum_{k=1}^{K} \beta_k x_{ki} = E(y_i^*) \quad (4)$$

Here comes an essential assumption of proportional odds models: The proportional odds assumption. This assumption requires that every coefficient beta in equation (4) needs to be the same. In other words, the regression lines of i equations in (4) must be parallel. This is why this assumption is also commonly known as the parallel-lines assumption. (Williams, 2006)

Continuing with the expression, with each particular value of the observed y (In this case, y takes three values: 1,2 and 3.), equation (4) can be changed to the probability equations below:

$$P(y_i > j) = \frac{e^{x_i\beta - k_j}}{1 + e^{x_i\beta - k_j}};\ j = 1, 2, 3, \ldots, M - 1 \quad (5)$$

The equation (5) can be expressed as:

$$P(y_i = 1) = 1 - \frac{e^{x_i\beta - k_j}}{1 + e^{x_i\beta - k_j}}$$

$$P(y_i = j) = \frac{e^{x_i\beta - k_{j-1}}}{1 + e^{x_i\beta - k_{j-1}}} - \frac{e^{x_i\beta - k_j}}{1 + e^{x_i\beta - k_j}};\ j = 2, 3, \ldots, M - 1$$

$$P(y_i = M) = \frac{e^{x_i\beta - k_{M-1}}}{1 + e^{x_i\beta - k_{M-1}}}$$

Since we have three values of observed dependent variables (M=3) and two cut-off points (M-1 = 2) in this case, the above equations will be shortened to these three probability equations:

$$\begin{aligned} P(y_i = 1) &= \frac{1}{1 + e^{Z_i - k_1}} \\ P(y_i = 2) &= \frac{1}{1 + e^{Z_i - k_2}} - P(\frac{1}{1 + e^{Z_i - k_1}}) \\ P(y_i = 3) &= 1 + \frac{1}{1 + e^{Z_i - k_2}} \end{aligned} \quad (6)$$

The interpretation of the Ordered Logit Model is not as straightforward as in the OLS method (Wooldridge, 2012). Moreover, Wooldridge also suggests that regression results expressed as exponentiated coefficients and log-odd ratios only tell us the relationship signs of each explanatory variable but not the magnitude of these differences.

Besides, when regression results are expressed as marginal effects (In other words, adjusted predictions), like the expressions in equations (6), the interpretations will be much more meaningful and visible. With marginal effects, we can define values for each explanatory variable in our regression model and then calculate the probability that the event will happen for a specific individual with the chosen values (Williams, 2012). Hence, this paper will also provide three regression results tables: The result table with exponentiated coefficients, the log-odd ratios, and the marginal effects (At mean values for continuous variables and at threshold values for factors variables) to make the results more tangible for further interpretations.

4.2 Data source and data descriptions

This research paper will utilize the Vietnam Household Living Standards Surveys (VHLSS) in 2014, 2016, and 2018. Every VHLSS survey is conducted once every two years by the General Statistics Office, under the supervision of the Ministry of Planning and Investment of Vietnam and the World Bank's technical assistance. This survey aims to observe the socioeconomic changes based on Vietnamese households' living conditions. In the analysis, we will mainly focus on the data in the 3 Sections below in the survey's results:

Section 4: Income of households.

Section 5: Expenditure of households.

Section 8: Participating in aid schemes status of households.

Based on the proposed econometric approach in Section 4.1, we will regress three equations for 2018, 2016, and 2014 respectively. The summary statistic of variables will be presented in Table 4.1 below:

Table 4.1

A Summary of the Descriptive Statistics

2018	Obs.	Mean	Std. Dev.	Min	Max
FI LEVEL					
1	34,437	.89	.32	0	1
2	34,437	.11	.31	0	1
3	34,437	.01	.08	0	1
Age	34,437	52.44	13.95	12	113
Age^2	34,437	2,944.12	1,560.22	144	12,769
Gender					
Female	34,437	.23	.42	0	1
Male	34,437	.77	.42	0	1
Marriage status					
Other	34,437	.47	.5	0	1
Married	34,437	.53	.5	0	1
Employed or not					
Unemployed	34,437	.16	.36	0	1
Employed	34,437	.84	.36	0	1
Educational level					
Did not complete high school	34,437	.76	.43	0	1
High school	34,437	.16	.37	0	1
College/University	34,437	.08	.27	0	1
Urban or rural area					
Rural	34,437	.69	.46	0	1
Urban	34,437	.31	.46	0	1
Relative income	34,437	1	.8	.01	25.46
Log (Total income per household)	34,437	11.56	.87	6.21	15.85
Nearest bank branch distance	34,437	9.11	3.73	4.3	14.3
2016	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
FI LEVEL					
1	43,459	.91	.28	0	1
2	43,459	.08	.27	0	1
3	43,459	.01	.07	0	1
Age	43,459	51.56	13.67	14	111
Age^2	43,459	2,845.69	1,508.17	196	12,321

Gender					
Female	43,459	.24	.43	0	1
Male	43,459	.76	.43	0	1
Marriage status					
Other	43,459	.19	.39	0	1
Married	43,459	.81	.39	0	1
Employed or not					
Unemployed	43,459	.16	.36	0	1
Employed	43,459	.84	.36	0	1
Educational level					
Did not complete high school	43,459	.77	.42	0	1
High school	43,459	.15	.36	0	1
College/University	43,459	.08	.27	0	1
Urban or rural area					
Rural	43,459	.69	.46	0	1
Urban	43,459	.31	.46	0	1
Relative income	43,459	1.01	.89	0	41.6
Log (Total income per household)	43,459	11.38	.87	3.91	15.74
Nearest bank branch distance	43,459	10.01	3.51	5.3	15.5
2014	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
FI LEVEL					
1	7,959	.94	.23	0	1
2	7,959	.06	.23	0	1
3	7,959	0	.06	0	1
Age	7,959	51.14	13.62	16	105
Age^2	7,959	2,801.16	1,506.69	256	11,025
Gender					
Female	7,959	.25	.44	0	1
Male	7,959	.75	.44	0	1
Marriage status					
Other	7,959	.2	.4	0	1
Married	7,959	.8	.4	0	1
Employed or not					
Unemployed	7,959	.34	.47	0	1
Employed	7,959	.66	.47	0	1
Educational level					
Did not complete high school	7,959	.77	.42	0	1

High school	7,959	.16	.37	0	1
College/University	7,959	.07	.26	0	1
Urban or rural area					
Rural	7,959	.69	.46	0	1
Urban	7,959	.31	.46	0	1
Relative income	7,959	1.01	.73	0	12.72
Log (Total income per household)	7,959	11.28	.84	5.52	15.21
Nearest bank branch distance	7,959	9.87	3.18	5.3	14.7

Source: Author's calculations using Stata

4.3 Variables descriptions

4.3.1 Financial Inclusion level (Instead of Financial Inclusion index)

In this study, following the method of Nguyen et al. (2021), we also plan to use a similar set of eight questions in Section 8 of VHLSS datasets to construct the IFI, but only for the household level. Besides, we also remove the last question (m8c5) because of the unclear intention of this question. Since households can borrow money or goods from their relatives or neighbors, the "m8c5" question cannot represent this paper's expected degree of financial inclusion.

The entire question descriptions of Section 8 can be listed in Table 4.2 below:

Table 4.2

Question descriptions of Section 8

Variable	Question	Type	Section in VHLSSs
m8c3a	Has your household got a bank account at this moment?	Dummy (1 = Yes)	Section 8
m8c3b	Has your household got a saving book at this moment?	Dummy (1 = Yes)	Section 8
m8c4a	Has your household used an ATM (Debit) card at this moment?	Dummy (1 = Yes)	Section 8
m8c4b	Has your household used a credit card at this moment?	Dummy (1 = Yes)	Section 8
m8c4c	Has your household got any life insurance at this moment?	Dummy (1 = Yes)	Section 8
m8c4d	Has your household got any non-life insurance at this moment?	Dummy (1 = Yes)	Section 8
m8c4e	Has your household got any stock or securities at this moment?	Dummy (1 = Yes)	Section 8
m8c5	Has your household borrowed any money or goods over the last 12 months?	Dummy (1 = Yes)	Section 8

Source: VHLSS datasets from 2014 to 2018

The original approach to calculating the local financial inclusion index of Nguyen et al. (2021) is simple and effective. First, for each household in a particular province, they aggregate all of these eight questions. Next, they sum all of the scores by province. Finally, they divided it by the product of the total number of households in this province and the total number of questions above. This approach ensures that each provincial index will range between zero and one (i.e., normalizing process). However, in this study, our purpose is to construct a household-level index of financial inclusion. Hence, in the first step, we aggregate the results of all the above questions (Excluding the "m8c5" question) into one single financial inclusion index for each household:

$$\boldsymbol{IFI_{i,t}} = m8c3a_{i,t} + m8c3b_{i,t} + m8c4a_{i,t} + m8c4b_{i,t} + m8c4c_{i,t} + m8c4d_{i,t} + m8c4e_{i,t}$$

Where:

$IFI_{I,t}$ denotes the financial inclusion index of household i in year t

$M8c3a_{i;t}$; $M8c3b_{i;t}$; $M8c4a_{i;t}$; $M8c4b_{i;t}$; $M8c4c_{i;t}$; $M8c4d_{i;t}$; $M8c4e_{i;t}$ are the Yes/No answers to the questions of the household i in year t (Takes the value 0 or 1 only)

Next, each IFI will be categorized into three levels of financial inclusion (LFI), as shown in Table 4.3 below:

Table 4.3

Categorizing the level of financial inclusion

Financial inclusion index (LFI)	Level of financial inclusion (LFI)
From 0 to 2	Low level – Takes the value of 1
From 3 to 4	Medium level – Takes the value of 2
From 5 to 7	High level – Takes the value of 3

Source: Author's allocations

4.3.2 Total income of each household

In the original VHLSS data, no question directly asks about the household's total income. Therefore, we manually calculate this variable based on some selected questions in Section 4: Income. This section divided the household's income into five main parts: Part 4A - Income from wages and salaries, part 4B – Income from agriculture activities, part 4C – Income from business, production, and other non-agriculture activities, and part 4D – Other revenues. At the end of each part, they have a question that sums up all the revenue and cost, which we utilize to calculate the whole income level for this paper. Based on the mentioned method, the total income level for each household will be calculated by the below function:

$$TI = m4atn + (m4b1t - m4b1c) + (m4b21t - m4b21c) + (m4b22t - m4b22c) + (m4b3t - m4b3c) + (m4b4t - m4b4c) + (m4b5t - m4b5c) + (m4ct - m4cc) + m4dtn$$

Where:

TI denotes the total income of household i in year t

m4atn is the sum of income from wages and salaries

m4b1t is the total revenue from harvested crops

m4b1c is the total cost of harvested crops

m4b21t is the total revenue from livestock farming

m4b21c is the total cost of livestock farming

m4b22t is the total revenue from hunting activities

m4b22c is the total cost of hunting activities

m4b3t is the total revenue from agriculture services

m4b3c is the total cost of agriculture services

m4b4t is the total revenue from forestry

m4b4c is the total cost of forestry

m4b5t is the total revenue from aquaculture

m4b5c is the total cost of aquaculture

4b5t is the total revenue from the business, production, and other non-agriculture activities

m4b5c is the total cost of business, production, and other non-agriculture activities

m4b5t is the total income from other revenues

4.3.3 Household's "relative income" index

According to Binh et al. (2021) and Li (2018), the relative income (Or distance income) is the difference between the average income of the area that the household currently lives in and the total income of this household. Binh et al. (2021) used the Vietnam Access to Resources Household Survey (VARHS) to calculate this variable in the original paper. Motivating by this approach, but with changes in calculating the average income of all household members at the district level, we construct the household relative income (HRI) variable with the VHLSS data as below:

$$\text{HRI}_{i,t} = \frac{TI_{i,t}}{AIP_t}$$

Where:

$TI_{i,t}$ denotes the total income of all members in household i in year t

$AIPI_t$ represents the average income of the particular district that this household currently lives in

4.3.4 Distance to the nearest commercial bank

For the distance from the nearest bank variable, as this information does not have in the questionnaires of VHLSS datasets, the data of this variable will be manually gathered using the VHLSS handbook of the General Statistics Office (GSO), which is also known as "The results of the VHLSS 2018 handbook". In Section 11.21: "Access to the infrastructure by region," the authors provide the average distance from households to their nearest financial institution at the provincial level. In this research paper, the distance from the nearest official bank branch will be used as one of the leading indicators of financial inclusion. The distance data from this handbook will be described in Table 4.4 below:

Table 4.4

Distance to the nearest official bank branch

Region	Distance		
	2014	**2016**	**2018**
Central Highlands	**14.5**	**15.2**	**13.8**
Province - Đắk Lắk	14.5	15.2	13.8
Province - Đắk Nông	14.5	15.2	13.8
Province - Gia Lai	14.5	15.2	13.8
Province - Kon Tum	14.5	15.2	13.8
Province - Lâm Đồng	14.5	15.2	13.8
Mekong River Delta	**10.1**	**9.2**	**9.7**
City - Cần Thơ	10.1	9.2	9.7
Province - An Giang	10.1	9.2	9.7
Province - Bạc Liêu	10.1	9.2	9.7
Province - Bến Tre	10.1	9.2	9.7
Province - Cà Mau	10.1	9.2	9.7
Province - Đồng Tháp	10.1	9.2	9.7
Province - Hậu Giang	10.1	9.2	9.7
Province - Kiên Giang	10.1	9.2	9.7
Province - Long An	10.1	9.2	9.7
Province - Sóc Trăng	10.1	9.2	9.7
Province - Tiền Giang	10.1	9.2	9.7
Province - Trà Vinh	10.1	9.2	9.7
Province - Vĩnh Long	10.1	9.2	9.7
North Central and Central coastal areas	**10.1**	**10.7**	**10.7**
City - Đà Nẵng	10.1	10.7	10.7
Province - Bình Định	10.1	10.7	10.7
Province - Bình Thuận	10.1	10.7	10.7
Province - Hà Tĩnh	10.1	10.7	10.7
Province - Khánh Hoà	10.1	10.7	10.7
Province - Nghệ An	10.1	10.7	10.7
Province - Ninh Thuận	10.1	10.7	10.7
Province - Phú Yên	10.1	10.7	10.7
Province - Quảng Bình	10.1	10.7	10.7
Province - Quảng Nam	10.1	10.7	10.7
Province - Quảng Ngãi	10.1	10.7	10.7
Province - Quảng Trị	10.1	10.7	10.7
Province - Thanh Hoá	10.1	10.7	10.7
Province - Thừa Thiên Huế	10.1	10.7	10.7
Northern midlands and mountain areas	**14.7**	**15.5**	**15.5**
Province - Bắc Giang	14.7	15.5	15.5
Province - Bắc Kạn	14.7	15.5	15.5
Province - Cao Bằng	14.7	15.5	15.5

Province - Điện Biên	14.7	15.5	15.5
Province - Hà Giang	14.7	15.5	15.5
Province - Hoà Bình	14.7	15.5	15.5
Province - Lai Châu	14.7	15.5	15.5
Province - Lạng Sơn	14.7	15.5	15.5
Province - Lào Cai	14.7	15.5	15.5
Province - Phú Thọ	14.7	15.5	15.5
Province - Sơn La	14.7	15.5	15.5
Province - Thái Nguyên	14.7	15.5	15.5
Province - Tuyên Quang	14.7	15.5	15.5
Province - Yên Bái	14.7	15.5	15.5
Red River Delta	**5.3**	**5.3**	**5.2**
City - Hà Nội	5.3	5.3	5.2
City - Hải Phòng	5.3	5.3	5.2
Province - Bắc Ninh	5.3	5.3	5.2
Province - Hà Nam	5.3	5.3	5.2
Province - Hải Dương	5.3	5.3	5.2
Province - Hưng Yên	5.3	5.3	5.2
Province - Nam Định	5.3	5.3	5.2
Province - Ninh Bình	5.3	5.3	5.2
Province - Quảng Ninh	5.3	5.3	5.2
Province - Thái Bình	5.3	5.3	5.2
Province - Vĩnh Phúc	5.3	5.3	5.2
South East	**8.4**	**8.1**	**8.1**
City - Hồ Chí Minh	8.4	8.1	8.1
Province - Bà Rịa - Vũng Tàu	8.4	8.1	8.1
Province - Bình Dương	8.4	8.1	8.1
Province - Bình Phước	8.4	8.1	8.1
Province - Đồng Nai	8.4	8.1	8.1
Province - Tây Ninh	8.4	8.1	8.1

Source: VHLSS 2018 Handbook

4.3.5 Other control variables

As mentioned above, together with the expected determinants of financial inclusion, some control variables about households' characteristics, such as age, gender, marriage status, employment status, educational level, as well as the living area of households' head, will also be employed in the model based on previous literature, to examine the effect of each of these factors on the level of financial inclusion of households.

Chapter 5
Results and interpretations

5.1 Maps of financial inclusion and its determinants

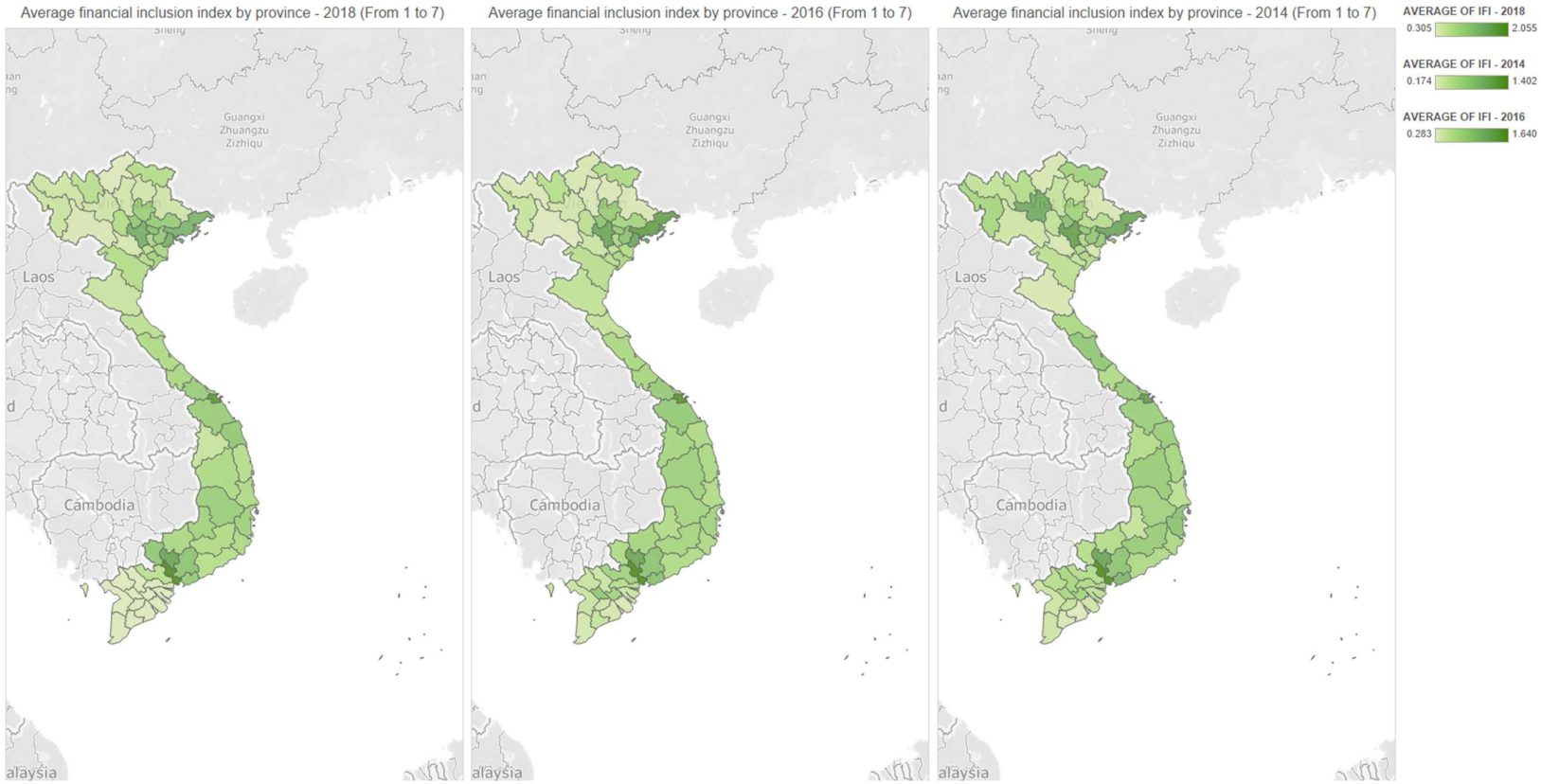
In this section, to capture the possible differences in the level of financial inclusion and its determinants (households' income and relative income index) across provinces, as well as to predict the sign of relationship among these three variables, maps of average provincial financial inclusion level, households' income, and their relative income index are illustrated in Figure 5.1, Figure 5.2 and Figure 5.3, respectively.

As exhibited in Figure 5.1 and Figure 5.2, in both three years from 2014 to 2018, we can clearly observe that the distribution of financial inclusion across provinces is heavily unbalanced. In some urban cities in North and South regions (such as Ha Noi and Ho Chi Minh City), the level of financial inclusion is significantly higher than in other areas, in line with the study of Nhan et al. (2021) and Tuyen & Van (2021) (also see Appendix 2). The provincial distribution of households' income level follows the same trend since households living in central provinces tend to gain more income than other provinces. The first-step perception indicates the indispensable role of financial inclusion in raising households' income, especially in key economic cities and regions. Interestingly, the visualizations from Figure 5.3 reveals the opposite tendency of the households' relative income variable. Since the relative income of center areas in North and South regions is smaller than in non-center areas, the inequality issue in terms of income distribution across households seems to correlate with financial inclusion negatively.

Together with these first-step predictions using visualized map charts, the empirical regression results will be presented in detail in Section 5.2 and Section 5.3 below.

Figure 5.1

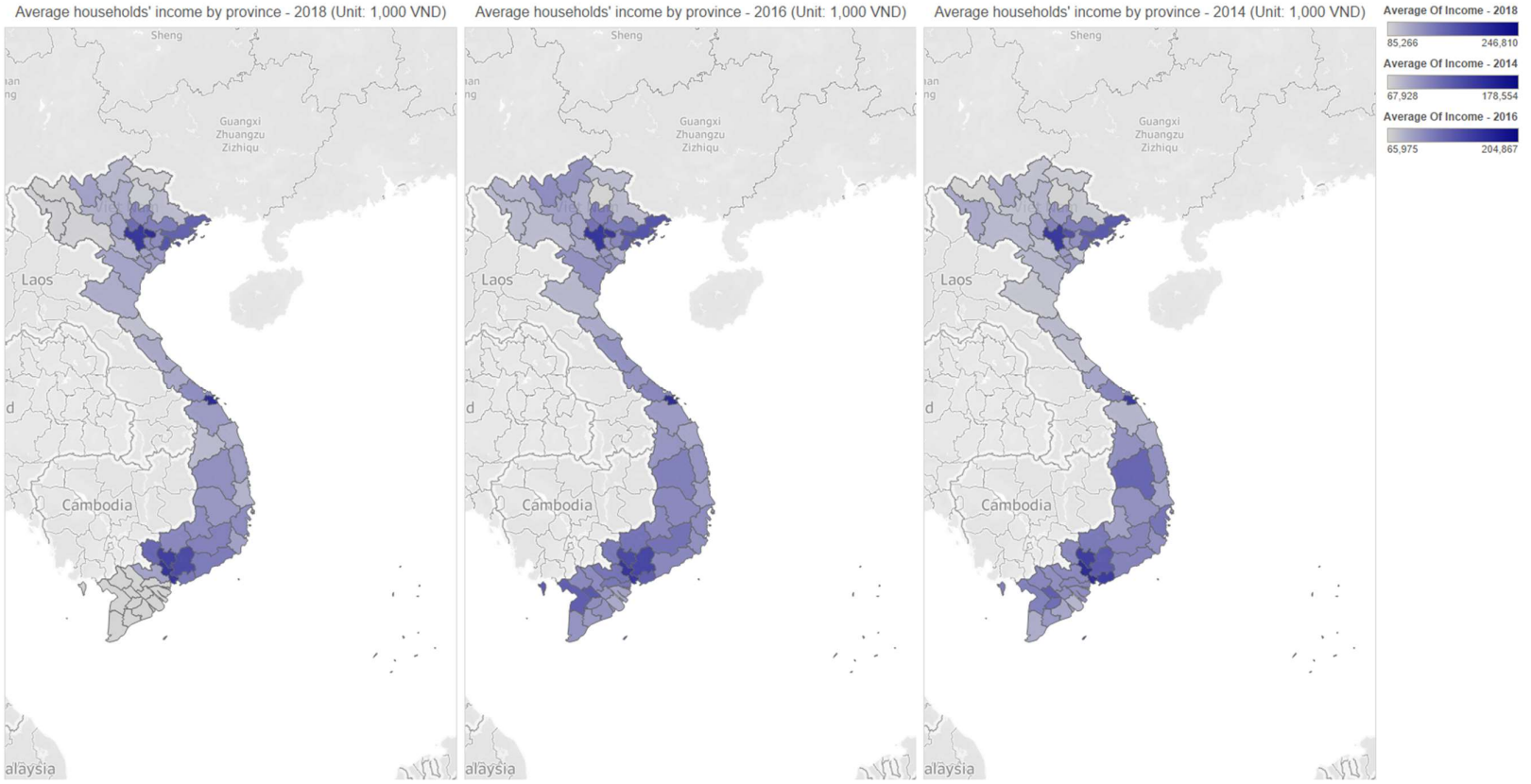
The average value of financial inclusion index by province (From 1 to 7)



Source: Author's visualizations using Tableau

Figure 5.2

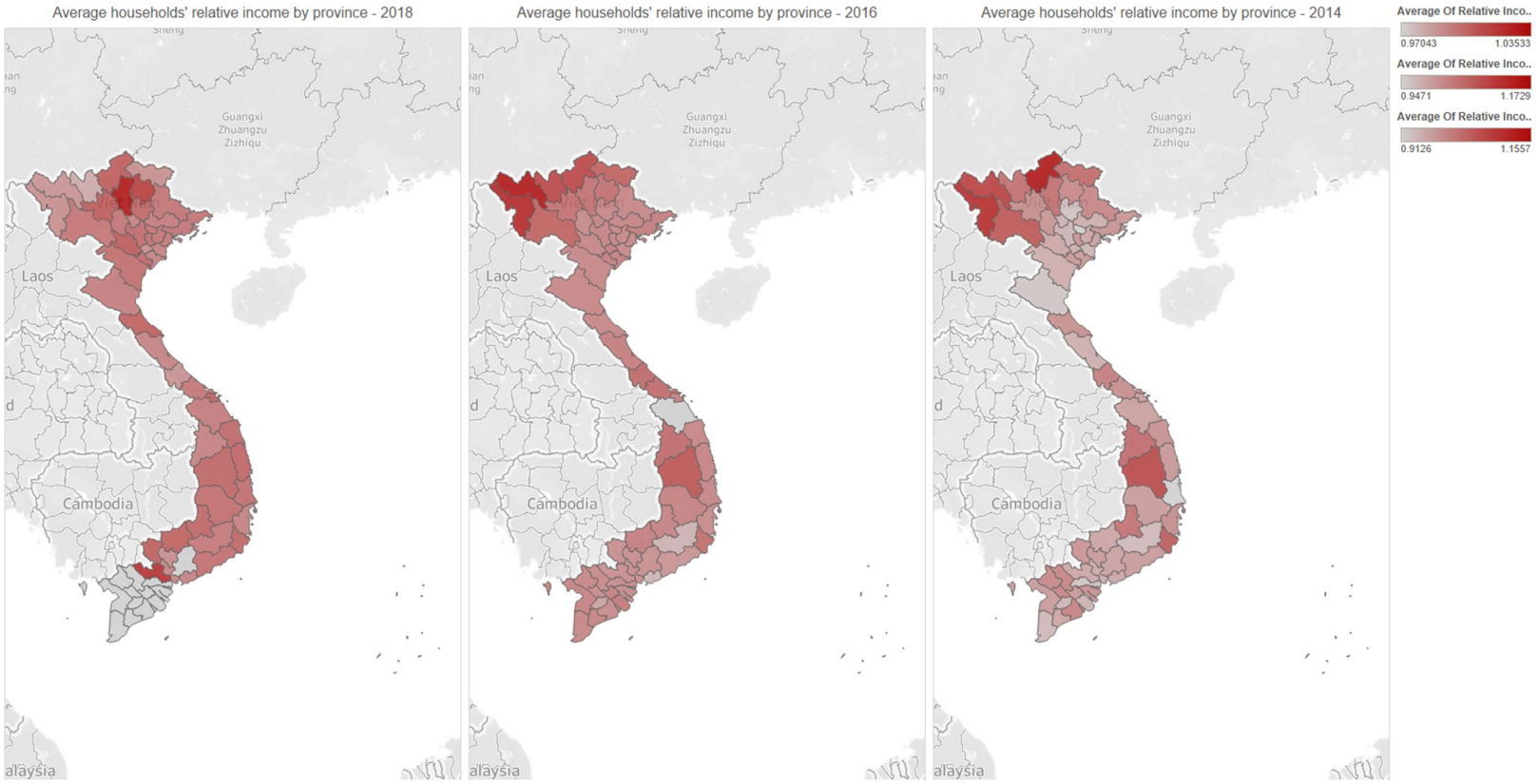
The average value of households' income level by province (Unit: 1,000 VND)



Source: Author's visualizations using Tableau

Figure 5.3

The average value of households' relative income level by province (Unit: 1,000 VND)



Source: Author's visualizations using Tableau

5.2 The regression results using the Ordinary Least Squares

5.2.1 Reasons for running OLS and its post-estimation tests

Generally, with discrete and ordinal dependent variables like the case of the level of financial inclusion, as we currently employ in this paper, the Ordinary Least Square method might be biased since it only treats the dependent variables as continuous values. Therefore, the (Partial) proportional-odd model (Or Ordered logit model) is the most appropriate approach to examine the relationship between the level of financial inclusion and other explanatory variables (Wooldridge, 2010).

Having said that, it is generally accepted that the linear model calculated using OLS can still provide an acceptable estimation of the real partial effects of independent variables (Wooldridge, 2012). Hence, we would also like to use Ordinary Least Squares (OLS) in this paper as a first-step regression to capture the signs of coefficients of each regressor, using the financial inclusion index (IFI) as the dependent variable. As mentioned in Section 4.3, unlike the level of financial inclusion (LFI), which only takes three values from 1 to 3, the IFI takes eight values from 0 to 7, reducing possible biases from OLS regression. After that, we can compare the regression results from OLS and the ordered logit model to get more meaningful interpretations. Another beneficial aspect of running OLS at the first step is that we can quickly run some after-regression diagnostics tests (Post tests) to detect possible and conspicuous problems such as heteroskedasticity and multicollinearity.

As shown in Table 5.1, the Breusch-Pagan (BP) test indicates that all OLS regression results in 2018, 2016, and 2014 suffered from the issue of heteroskedasticity, which can violate one of the crucial assumptions of OLS, which requires the error terms have constant variances for observed samples. White suggested that with large enough sample sizes, the heteroskedasticity issue in OLS can be corrected using robust standard errors (Gujarati, 2011). The OLS regression results with robust standard errors will be provided in Table 5.4 below to overcome this problem.

Together with the heteroskedasticity issue, the (imperfect) multicollinearity is also one of the common issues when we aim to plug a considerable number of explanatory variables into the model. If we find any sign of imperfect multicollinearity between any pair of variables in the model, the OLS estimation might still be unbiased, but it will cause problems in providing reasonable variances. In other words, the variances of our estimated coefficients will be inflated. As a result, the significance level of regressors' coefficients will be inaccurate (Gujarati, 2011). To detect multicollinearity, the Variance Inflating Factor (VIF) and Pearson's pairwise correlation test (Pwcorr), two of the most common diagnostics test for this issue, will be employed. Table 5.2 and Table 5.3 below show the results of the VIF test and Pwcorr test, respectively.

Several rules of thumb are mentioned about the upper limit of the VIF values. Gujarati (2011) suggests that the VIF values below 2 are appropriate, while Wooldridge (2012) relaxed this threshold since every VIF values below 10 are not a problem. Looking at the results of the VIF test in Table 5.2, we can confidently conclude that there is not any severe multicollinearity here since every single VIF value of each regressor is lower than 10 (Except for two variables: Age and Age^2, because these two variables are correlated by nature since Age^2 is the squared term of Age).

Different results are found in Pearson's pairwise correlation test. As Alan (2018) stated, the general rule of thumb for deciding the correlation level between two independent variables is that: If the absolute value of the computed correlation number is between 0.1 and 0.3,

there is a weak correlation between two variables; if the absolute value of the computed correlation number is between 0.3 and 0.5, there is an average correlation between two variables. Any absolute value of the correlation number higher than 0.5 indicates a high correlation between these two variables. Looking at Table 5.3, Pearson's pairwise correlation test does not detect any noticeable multicollinearity between pairs of independent variables, except for two pairs: Age versus Age^2 and Total Income Per Household versus Relative Income. Since Relative Income is the variable calculated by dividing Total Income Per Household by the average income of the particular district, this relationship will be accepted to correlate by nature.

5.2.2 Estimation results from OLS with robust standard errors

Looking at the regression results from OLS with robust standard errors in Table 5.4, we can draw several first-step interpretations.

First, we can see the coefficients of the Log(Total Income) variable in both three years are positive and significant at 1% level. Besides, these coefficients of Log(Total Income) (Log(TI)) seem to increase year-over-year, from 0.0374 in 2014 to 0.0453 in 2016 and 0.0896 in 2018, which suggests that the effect of raising households' income on enhancing financial inclusion is getting higher over time. As the marginal effect and elasticity of the log term variable in our linear model are very straightforward, to estimate the change in the level of IFI when the total income increases by 1 percent, we can simply multiply the coefficient by the exact value of the level of financial inclusion that we want to interpret. For instance, using the coefficient of Log(TI) = 0.0896 in 2016, the elasticity at level 2 of IFI will be equal: $0.0896 \frac{1}{IFI(=2)} = 0.0448$, which indicates that when the total income increases by 1 percent, the IFI will increase by 0.0448%. This calculation can give us the overall effect of income on financial inclusion. However, as mentioned above, this kind of estimation will be biased since IFI is a discrete dependent variable and should not be treated as continuous.

Second, the coefficients of household relative income variable (HRI) in 2018 (0.0315) and 2014 (0.0338) are significantly and negatively correlated with the IFI at 1% and 5% levels, respectively. However, since the coefficient in 2016 (0.001) is not significant and relatively small compared to 2018 and 2014, we cannot draw the same conclusion as the Log(TI) variable that the relationships between HRI and IFI are not going on the same direction over time.

Third, the coefficients of the distance to the nearest public bank variable (Distance) and the living in urban areas variable (Location) are negatively correlated with IFI in both three regressions in 2018, 2014, and 2012 at 1% level, implying that households who live the urban areas and near commercial banks tend to have higher financial inclusion level. Similar to the total income variable, both three coefficients in 2018 (0.127), 2016 (0.0768) and 2014 (0.0664) of the Location variables also suggest that the positive effect of living in urban areas on increasing financial inclusion level tends to increase over time.

The coefficient signs of other explanatory variables also bring some interesting perspectives. However, since the main focus of this paper is to investigate the results of the Ordinal Logit Model, in Section 5.3 below, we will precisely interpret these variables using marginal effects at mean values and compare them with the results from OLS to yield more insights about the magnitude of these effects.

Table 5.1

Breusch-Pagan test for heteroskedasticity issues

	Breusch-Pagan test for heteroskedasticity		
H_0: Constant variance	**2018**	**2016**	**2014**
chi2(1)	15788.45	23470.48	6604.85
Prob > chi2	0.0000	0.0000	0.0000

Notes: A significant test statistic indicates that the null hypothesis H_0 has been rejected

Source: Author's calculations using Stata

Table 5.2

Variance Inflating Factor (VIF) test

	VIF test		
Variables	**VIF - 2018**	**VIF - 2016**	**VIF - 2014**
Age	37.16	39.08	40.06
Age^2	37.29	39.93	40.94
Gender	1.17	1.64	1.72
Marriage status	1.15	1.75	1.91
Employed	1.07	1.17	1.03
Graduated from high school	1.02	1.07	1.07
Graduated from college/university	1.02	1.12	1.13
Living in urban areas	1.21	1.28	1.26
Relative income	2.16	2.01	2.44
Log(Total Income)	2.49	2.35	2.99
Distance from the nearest commercial bank	1.09	1.09	1.10
Mean VIF	**7.89**	**8.41**	**8.70**

Source: Author's calculations using Stata

Table 5.3

Pearson's correlation test

2018	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Age	1.000									
(2) Age2	0.986*	1.000								
(3) Gender	-0.223*	-0.229*	1.000							
(4) MarriageStatus	-0.161*	-0.170*	0.343*	1.000						
(5) EmployedOrNot	-0.025*	-0.025*	0.009	-0.001	1.000					
(6) EducationalLevel	-0.016*	-0.016*	0.004	0.018*	-0.004	1.000				
(7) UrbanOrRuralArea	0.039*	0.040*	-0.015*	-0.014*	-0.161*	-0.001	1.000			
(8) RelativeIncome	-0.012*	-0.013*	0.008	0.004	0.113*	0.007	0.087*	1.000		
(9) LogTotalIncomePerHousehold	0.029*	0.028*	-0.021*	-0.009	0.094*	0.006	0.321*	0.706*	1.000	
(10) NearestBankBranchDis	-0.073*	-0.068*	0.034*	0.013*	0.117*	0.001	-0.072*	0.002	-0.178*	1.000
2016	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
(1) Age	1.000									
(2) Age2	0.987*	1.000								
(3) Gender	-0.176*	-0.182*	1.000							
(4) MarriageStatus	-0.293*	-0.312*	0.610*	1.000						
(5) EmployedOrNot	-0.294*	-0.311*	0.120*	0.138*	1.000					
(6) EducationalLevel	-0.091*	-0.092*	-0.004	0.080*	-0.037*	1.000				
(7) UrbanOrRuralArea	0.049*	0.043*	-0.154*	-0.056*	-0.153*	0.295*	1.000			
(8) RelativeIncome	-0.048*	-0.059*	0.049*	0.087*	0.102*	0.128*	0.070*	1.000		
(9) LogTotalIncomePerHousehold	-0.067*	-0.088*	0.029*	0.124*	0.104*	0.239*	0.297*	0.686*	1.000	
(10) NearestBankBranchDis	-0.146*	-0.138*	0.048*	0.053*	0.109*	-0.043*	-0.067*	0.010*	-0.147*	1.000
2014	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
(1) Age	1.000									
(2) Age2	0.987*	1.000								
(3) Gender	-0.169*	-0.175*	1.000							
(4) MarriageStatus	-0.299*	-0.322*	0.629*	1.000						
(5) EmployedOrNot	-0.065*	-0.083*	0.046*	0.033*	1.000					
(6) EducationalLevel	0.026*	0.010	-0.053*	0.026*	0.009	1.000				
(7) UrbanOrRuralArea	0.027*	0.021	-0.153*	-0.035*	-0.105*	0.251*	1.000			
(8) RelativeIncome	-0.077*	-0.099*	0.099*	0.208*	0.011	0.216*	0.093*	1.000		
(9) LogTotalIncomePerHousehold	-0.128*	-0.159*	0.098*	0.257*	-0.004	0.314*	0.318*	0.740*	1.000	
(10) NearestBankBranchDis	-0.146*	-0.139*	0.039*	0.054*	0.045*	-0.063*	-0.061*	0.022*	-0.143*	1.000

*Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Source: Author's calculations using Stata

Table 5.4

The Ordinary Least Squares regression – With robust standard errors

	2018		2016		2014	
	Coefficients	**t-statistics**	**Coefficients**	**t-statistics**	**Coefficients**	**t-statistics**
Age	-0.00102	(-1.34)	0.00453***	(8.14)	-0.0000395	(-0.04)
Age^2	0.00000949	(1.38)	-0.0000373***	(-7.32)	0.00000357	(0.36)
Gender	-0.00457	(-1.01)	-0.0287***	(-6.05)	-0.0220*	(-2.22)
Marriage status	0.00297	(0.80)	0.0350***	(7.42)	0.0116	(1.17)
Employed	-0.0448***	(-8.42)	-0.0295***	(-6.49)	-0.00369	(-0.62)
Graduated from high school	0.00523	(1.12)	0.0802***	(16.95)	0.0222*	(2.52)
Graduated from college/university	0.0166*	(2.56)	0.262***	(27.99)	0.148***	(7.65)
Living in urban areas	0.127***	(27.61)	0.0768***	(20.91)	0.0664***	(9.16)
Relative income	0.0315***	(6.43)	0.00100	(0.36)	0.0338**	(3.03)
Log(Total Income)	0.0896***	(26.66)	0.0453***	(18.42)	0.0374***	(5.35)
Distance (Kilometer)	-0.00343***	(-7.64)	-0.00144***	(-3.50)	-0.00221*	(-2.48)
Constant	0.107**	(2.64)	0.423***	(14.63)	0.595***	(7.73)
N	34437		43459		7959	
R-Squared	0.146		0.136		0.122	
Adjusted R-Squared	0.146		0.136		0.121	

Notes: t statistics in parentheses

** p < 0.05, ** p < 0.01, *** p < 0.001*

Source: Author's calculations using Stata

5.3 The regression results using Ordered logit regression

Using the well-known "ologit" command from Stata, the ordered logit regression (OLM) results in 2018, 2016, and 2014 will be provided in this section. To better capture the tangible and detailed effect of each determinant on financial inclusion, the regression results will be expressed as exponentiated coefficients, log-odds values, and marginal effects (At mean values for continuous variables and at threshold values for factors variables) in Table 5.5, Table 5.6 and Table 5.9 respectively.

Since directly interpreting the coefficients in Table 5.5 and Table 5.6 is quite tricky, we plan to divide the interpretation into three parts: In the first part, we capture the effect signs of each determinant in both OLM and OLS to see whether we have any difference here between two methods in terms of coefficient signs. In the second part, the Brant test will be employed with explanations to check whether the parallel-lines assumption holds. Finally, the third part will provide detailed interpretations of each determinant based on the results of the marginal effects.

5.3.1 The coefficients signs of OLM compared to OLS

Based on the OLM regression results in Table 5.5, we summarize the (exponential) coefficients of this regression and the coefficients of the OLS regression with robust standard errors in Table 5.4 above, together with the expected sign of each variable based on the review of literature in Chapter 2.

In general, the signs of OLM and OLS's coefficients are the same across three regressions in 2018, 2016, and 2014; except for Age, Age^2, Gender, and Relative income. Besides, the expected signs of four variables: Graduated from high school, Graduated from college/university, Living in urban areas, Relative income, Log(Total Income), and Distance is not different from our regression results, implying that the empirical results of these determinants will be supported by previous literature. Moreover, the different signs of relative income coefficients between OLM and OLS might suggest that the OLM method yields better estimated and unbiased results than OLS since these negative relationships from OLM results are well supported by the previous studies. (Prete, 2013; Park & Mercado, 2015; Zhang & Alberto, 2017). Unexpectedly, though there is limited literature about the relationship between employment status and financial inclusion, the negative signs across years of employed variables in both OLM and OLS results are also an interesting point to look at since it contradicts the common sense that employed individuals or in our case, employed household heads, tend to get better access to financial services and banking system, therefore have higher levels of financial inclusion.

Table 5.5

The Ordered Logit Model regression results – Exponentiated coefficients

	2018		2016		2014	
	Coefficients	**t-statistics**	**Coefficients**	**t-statistics**	**Coefficients**	**t-statistics**
Level of FI						
Age	-0.0153	(-1.89)	0.0864***	(8.44)	0.0216	(0.76)
Age^2	0.000125	(1.73)	-0.000750***	(-7.88)	-0.000201	(-0.76)
Gender	0.000848	(0.02)	-0.271***	(-5.60)	-0.0862	(-0.64)
Marriage status	0.00139	(0.03)	0.383***	(6.14)	0.0692	(0.37)
Employed	-0.303***	(-6.15)	-0.303***	(-5.98)	-0.0613	(-0.55)
Graduated from high school	0.00666	(0.13)	0.926***	(20.36)	0.457***	(3.49)
Graduated from college/ university	0.0775	(1.14)	1.717***	(34.67)	0.949***	(6.96)
Living in urban areas	0.840***	(20.30)	0.779***	(17.98)	0.800***	(6.50)
Relative income	-0.283***	(-9.84)	-0.176***	(-6.28)	-0.299***	(-3.49)
Log(Total Income)	2.016***	(40.18)	1.012***	(23.47)	2.105***	(14.29)
Distance (Kilometer)	-0.0180**	(-3.16)	-0.0224***	(-3.94)	-0.00324	(-0.18)
cut1	25.40***	(40.38)	16.83***	(30.27)	28.28***	(15.59)
cut2	28.80***	(45.16)	19.93***	(35.42)	31.53***	(17.18)
N	34437		43459		7959	
Chi-squared	5867.6		5279.6		958.4	
Degrees of Freedom	11		11		11	
p	0		0		1.74e-198	
Pseudo R-Squared	0.226		0.195		0.256	

Notes: t statistics in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001
Source: Author's calculations using Stata

Table 5.6

The Ordered Logit Model regression results – Log-odds values

	2018		2016		2014	
	Coefficients	**t-statistics**	**Coefficients**	**t-statistics**	**Coefficients**	**t-statistics**
Level of FI						
Age	0.985	(-1.89)	1.090***	(8.44)	1.022	(0.76)
Age^2	1.000	(1.73)	0.999***	(-7.88)	1.000	(-0.76)
Gender	1.001	(0.02)	0.763***	(-5.60)	0.917	(-0.64)
Marriage status	1.001	(0.03)	1.467***	(6.14)	1.072	(0.37)
Employed	0.738***	(-6.15)	0.739***	(-5.98)	0.941	(-0.55)
Graduated from high school	1.007	(0.13)	2.526***	(20.36)	1.580***	(3.49)
Graduated from college/ university	1.081	(1.14)	5.569***	(34.67)	2.582***	(6.96)
Living in urban areas	2.317***	(20.30)	2.179***	(17.98)	2.226***	(6.50)
Relative income	0.754***	(-9.84)	0.839***	(-6.28)	0.742***	(-3.49)
Log(Total Income)	7.511***	(40.18)	2.751***	(23.47)	8.207***	(14.29)
Distance (Kilometer)	0.982**	(-3.16)	0.978***	(-3.94)	0.997	(-0.18)
cut1	1.07636e+11***	(40.38)	20377134.3***	(30.27)	1.92011e+12***	(15.59)
cut2	3.21547e+12***	(45.16)	450692143.4***	(35.42)	4.94865e+13***	(17.18)
N	34437		43459		7959	
Chi-squared	5867.6		5279.6		958.4	
Degrees of Freedom	11		11		11	
p	0		0		1.74e-198	
Pseudo R-Squared	0.226		0.195		0.256	

Notes: t statistics in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Source: Author's calculations using Stata

Table 5.7
The summary of the coefficient signs – OLM versus OLS with robust standard errors

Variables	2018		2016		2014		Expected sign
	OLM coefficient	**OLS coefficient**	**OLM coefficient**	**OLS coefficient**	**OLM coefficient**	**OLS coefficient**	
Age	-0.0153	-0.00102	0.0864***	0.00453***	0.0216	-0.0000395	
Age^2	0.000125	0.00000949	-0.000750***	-0.0000373***	-0.000201	0.00000357	
Gender	0.000848	-0.00457	-0.271***	-0.0287***	-0.0862	-0.0220*	
Marriage status	0.00139	0.00297	0.383***	0.0350***	0.0692	0.0116	
Employed	-0.303***	-0.0448***	-0.303***	-0.0295***	-0.0613	-0.00369	
Graduated from high school	0.00666	0.00523	0.926***	0.0802***	0.457***	0.0222*	+
Graduated from college/university	0.0775	0.0166*	1.717***	0.262***	0.949***	0.148***	+
Living in urban areas	0.840***	0.127***	0.779***	0.0768***	0.800***	0.0664***	+
Relative income	-0.283***	0.0315***	-0.176***	0.00100	-0.299***	0.0338**	-
Log(Total Income)	2.016***	0.0896***	1.012***	0.0453***	2.105***	0.0374***	+
Distance (Kilometer)	-0.0180**	-0.00343***	-0.0224***	-0.00144***	-0.00324	-0.00221*	-

Source: Author's calculations using Stata

Table 5.8

Brant test for the parallel-lines assumption

	2018			**2016**			**2014**		
	chi2	**p>chi2**	**df**	**chi2**	**p>chi2**	**df**	**chi2**	**p>chi2**	**df**
All variables	**52.58**	**0.000**	**11**	**30.07**	**0.002**	**11**	**8.90**	**0.631**	**11**
Age	1.41	0.235	1	0.52	0.469	1	0.36	0.549	1
Age^2	1.45	0.229	1	0.59	0.443	1	0.34	0.559	1
Gender	3.02	0.082	1	1.50	0.221	1	0.12	0.733	1
Marriage status	1.88	0.171	1	0.34	0.560	1	0.00	0.985	1
Employed	1.25	0.264	1	2.54	0.111	1	0.10	0.749	1
Graduated from high school	0.00	1.000	1	1.49	0.222	1	1.11	0.291	1
Graduated from college/university	0.48	0.488	1	0.75	0.386	1	0.08	0.782	1
Living in urban areas	17.80	0.000	1	5.63	0.018	1	1.19	0.276	1
Relative income	0.02	0.876	1	1.99	0.158	1	1.75	0.186	1
Log(Total Income)	5.94	0.015	1	0.69	0.407	1	0.03	0.859	1
Distance from the nearest commercial state bank (Kilometer)	5.22	0.022	1	6.47	0.011	1	0.14	0.709	1

Notes: A significant test statistic provide evidence that the parallel regression assumption has been violated.

Source: Author's calculations using Stata

5.3.2 The Brant test for parallel-lines assumption

Williams (2006) mentioned that one of the most critical assumptions to be satisfied when employing proportional odds models is the proportional odds assumption. This assumption requires every βs in equation (4) in Section 3.1 to be the same. In other words, the regression lines of each equation in (4) must be parallel. This is why this assumption is also commonly known as the parallel-lines assumption. If this assumption can not be held, the regression results from OLM will be biased and inconsistent. Williams (2006) also suggests that one should consider switching the regression method from OLM to other methods, such as Generalized OLM or Multinominal Logit Model (MNL). Since we would like to observe the effects of our investigated determinants on the ordinal characteristic of LFI, the generalized OLM seems to be the optimal option. The Brant test, commonly known as the likelihood ratio test, will compare the Generalized OLM and OLM to see which method will be the most suitable one (In other words, this test checks the null hypothesis H_0 that whether the OLM is nested in the generalized OLM). Since the generalized OLM does not require every βs in equation (4) in Section 3.1 to be the same, the Brant test null hypothesis H_0 can be understood as the parallel-lines assumption being held.

The Brant test results in Table 5.8 indicate that the aggregated proportional odds assumption was violated in 2018 and 2016. The main issue comes from living in urban areas, and distance from the nearest public bank since the coefficients from these variables noticeably differ from other variables. As suggested by Williams (2006, 2016), to overcome this problem, we will employ the Generalized OLM test in two datasets in 2018 and 2016, using the "gologit2" command in Stata, with the "autofit" option. In short, the command with this option will re-run the Wald-test to check if any variable violates the assumption, nearly the same thing as the Brant test does, and after that, remove the restrictions of the parallel lines assumption only for violated variables. Using this command, we can confidently interpret the results without fear of violating this assumption.

5.3.3 The marginal effects interpretations of key determinants

As explained in Section 5.3.2 above, the average marginal effects using generalized OLM in 2018 and 2016; and OLM in 2014 will be provided in Table 5.9 below. Variables non-restricted to the parallel-lines assumption (When running the regressions with gologit2) will also be emboldened. Unlike the exponential coefficients and log-odds values, interpreting the average marginal effect is tangible and straightforward since it can provide the predicted probabilities at each value of LFI, depending on the changes of our independent variables. Looking at Table 5.9, we can see that, except for Age, Age^2, Gender, and Married status, all other variables do not change the signs of average marginal effects across years. Hence, for these variables, we can focus on interpreting the detailed effects in 2018 and compare the magnitude of these effects with 2016 and 2014 later.

First, take the total income's average marginal effects (AMEs) as an example for interpretation. The AMEs of the log(Total income) at LFI = 1 (-0.1611), LFI = 2 (0.1468), and LFI = 3 (0.0143) in 2018 imply that with other variables being held constant, on average, if the income of one particular household increases by one unit, the probability that the level of financial inclusion (LFI) of this household is about 16.11% less likely to be in the group of low LFI, 14.68% more likely to be in the group of average LFI and 1.43% more likely to be in the group of high LFI. The same signs of AMEs were also found in 2016 and 2014. This finding is also consistent with previous literature about the positive relationship between financial inclusion and income in general (Jauch & Watzka, 2011; Park & Mercado, 2015; Wang & Guan, 2016; Luan & Bauer, 2016; Zhang & Alberto, 2017)

Second, for the relative income variables (HRI), the AMEs of the HRI at LFI = 1 (0.0235), LFI = 2 (-0.0219), and LFI = 3 (-0.0016) in 2018 imply that with other variables being held constant, on average, if the relative income of one particular household increases by one unit, the probability that the level of financial inclusion (LFI) of this household is approximately 2.35% more likely to be in the group of low LFI, 2.19% less likely to be in the group of average LFI and 0.16% less likely to be in the group of high LFI. Similar to the vase of total income per household, this empirical result is in line with past studies. (Prete, 2013; Park & Mercado, 2015; Zhang & Alberto, 2017)

Third, the AME results of the distance from the nearest commercial bank are mixed across levels of financial inclusion. With other variables being held constant, on average, if the distance from the nearest commercial bank of one particular household increases by one unit, the probability that the level of financial inclusion (LFI) of this household is approximately 1.5% more likely to be in the group of low LFI, 1.7% less likely to be in the group of average LFI and 0.02% more likely to be in the group of high LFI. Although the distance from the financial institution is one of the main barriers that lower the growth level of financial inclusion (Sarma, 2008, 2012; Demirguc-Kunt & Klapper, 2012; Noelia & Tuesta, 2014; Svirydzenka, 2016), the effect of distance on the high level of financial inclusion is relatively small compared to other determinants.

The effect of other factors like Age, Age^2, Gender, Married, Employed, and Educational level should also be considered. The effect of Age and Age^2 is noticeably small in affecting the financial inclusion level, which is different from the findings of Luan & Bauer (2016). The effect of Gender seems to vary across the years, with only the result in 2016 being significant. The conclusion here is that being male acts as the factor that enhances financial inclusion only when the financial inclusion level of the household is low. Interestingly, the results also show that married and high-educated household heads positively affect financial inclusion levels only at medium and high levels. The location and distance seem to affect financial inclusion in the same manner. Precisely, households who live in urban areas have 7.02% less likely to be in the low financial inclusion level, 6.27% more likely to be in the average level, and 0.75% more likely to be in the high level.

Table 5.9

The average marginal effects (AMEs)

	2018		2016		2014	
	AME	**t-statistics**	**AME**	**t-statistics**	**AME**	**t-statistics**
Age						
Low LFI (LFI =1)	0.0012	(0.0007)	-0.0056***	(0.0007)	-0.0010	(0.0013)
Medium LFI (LFI =2)	-0.0011	(0.0006)	0.0052***	(0.0006)	0.0009	(0.0012)
High LFI (LFI =3)	-0.0001	(0.0000)	0.0004***	(0.0001)	0.0001	(0.0001)
Age^2						
Low LFI (LFI =1)	-0.0000	(0.0000)	0.0000***	(0.0000)	0.0000	(0.0000)
Medium LFI (LFI =2)	0.0000	(0.0000)	-0.0000***	(0.0000)	-0.0000	(0.0000)
High LFI (LFI =3)	0.0000	(0.0000)	-0.0000***	(0.0000)	-0.0000	(0.0000)
Gender						
Low LFI (LFI =1)	-0.0002	(0.0038)	0.0183***	(0.0034)	0.0039	(0.0062)
Medium LFI (LFI =2)	0.0002	(0.0036)	-0.0168***	(0.0031)	-0.0036	(0.0058)
High LFI (LFI =3)	0.0000	(0.0003)	-0.0015***	(0.0003)	-0.0003	(0.0005)
Marriage status						
Low LFI (LFI =1)	-0.0000	(0.0033)	-0.0230***	(0.0035)	-0.0030	(0.0081)
Medium LFI (LFI =2)	0.0000	(0.0030)	0.0213***	(0.0032)	0.0028	(0.0075)
High LFI (LFI =3)	0.0000	(0.0002)	0.0018***	(0.0003)	0.0002	(0.0006)
Employed						
Low LFI (LFI =1)	0.0256***	(0.0044)	0.0209***	(0.0037)	0.0028	(0.0050)
Medium LFI (LFI =2)	-0.0238***	(0.0041)	-0.0192***	(0.0034)	-0.0026	(0.0046)
High LFI (LFI =3)	-0.0018***	(0.0003)	-0.0017***	(0.0003)	-0.0002	(0.0004)
Graduated from high school						
Low LFI (LFI =1)	-0.0006	(0.0042)	-0.0642***	(0.0037)	-0.0204**	(0.0063)
Medium LFI (LFI =2)	0.0006	(0.0039)	0.0602***	(0.0035)	0.0192**	(0.0059)
High LFI (LFI =3)	0.0000	(0.0003)	0.0039***	(0.0003)	0.0013**	(0.0005)

Graduated from college/university						
Low LFI (LFI =1)	-0.0061	(0.0057)	-0.1576***	(0.0063)	-0.0502***	(0.0086)
Medium LFI (LFI =2)	0.0057	(0.0053)	0.1461***	(0.0058)	0.0468***	(0.0081)
High LFI (LFI =3)	0.0004	(0.0004)	0.0115***	(0.0009)	0.0034***	(0.0009)
Living in urban areas						
Low LFI (LFI =1)	**-0.0702*** **	(0.0036)	**-0.0525*** **	(0.0031)	-0.0363***	(0.0056)
Medium LFI (LFI =2)	**0.0627*** **	(0.0036)	**0.0475*** **	(0.0030)	0.0340***	(0.0053)
High LFI (LFI =3)	**0.0074*** **	(0.0008)	**0.0050*** **	(0.0007)	0.0023***	(0.0005)
Relative income						
Low LFI (LFI =1)	0.0235***	(0.0023)	**0.0119*** **	(0.0018)	0.0133***	(0.0038)
Medium LFI (LFI =2)	-0.0219***	(0.0021)	**-0.0115*** **	(0.0017)	-0.0124***	(0.0036)
High LFI (LFI =3)	-0.0016***	(0.0002)	**-0.0004**	(0.0002)	-0.0010**	(0.0003)
Log(Total income)						
Low LFI (LFI =1)	**-0.1611*** **	(0.0039)	-0.0660***	(0.0028)	-0.0941***	(0.0067)
Medium LFI (LFI =2)	**0.1468*** **	(0.0037)	0.0607***	(0.0026)	0.0873***	(0.0063)
High LFI (LFI =3)	**0.0143*** **	(0.0011)	0.0053***	(0.0004)	0.0068***	(0.0014)
Distance from the nearest commercial state bank						
Low LFI (LFI =1)	**0.0015*** **	(0.0005)	**0.0015*** **	(0.0004)	0.0001	(0.0008)
Medium LFI (LFI =2)	**-0.0017*** **	(0.0005)	**-0.0016*** **	(0.0004)	-0.0001	(0.0007)
High LFI (LFI =3)	**0.0002**	(0.0001)	**0.0001**	(0.0001)	-0.0000	(0.0001)
N	34437		43459		7959	

Notes: Standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Variables that are non-restricted to the parallel-lines assumption are emboldened

Source: Author's calculations using Stata

Chapter 6
Conclusions

Over the past decades, a significant body of literature has emphasized the role of financial inclusion in improving individuals' income, reducing poverty, and ameliorating income inequality. Nevertheless, together with the growing literature at the global level, in Vietnam, limited studies have been conducted to examine detailed characteristics of financial inclusion using micro-level datasets such as VHLSS or VARHS. Using the Vietnam Household Living Standard Survey (VHLSS) datasets in 2018, 2016, and 2014, this research paper examines the key determinants that influence financial inclusion in the Vietnamese context. Key findings are summarised as follows. First, the income level seems to lower the level of financial inclusion if this household has a low level of financial inclusion (IFI = 1) and has a positive impact on the level of financial inclusion if this household has a medium or high level of financial inclusion (IFI = 2 or 3). In short, the positive impact is dominant. This relationship is well-supported by various empirical findings. Second, the relative income has a positive impact on financial inclusion if this household has a low level of financial inclusion (IFI = 1); but negatively affects the level of financial inclusion if this household has a medium or high level of financial inclusion (IFI = 2 or 3). Differing from income, the negative relationship between relative income and financial inclusion is dominant. Finally, the location of households and the close distance from households to their nearest public bank plays a crucial role in improving financial inclusion, especially when the particular household is currently at the medium level of financial inclusion.

Policy implications have emerged based on the findings of this study. Policymakers can help increase individuals' income level and address income inequality by implementing policies supporting financial inclusion through various tools, such as easing the new credit/debit account opening procedure, increasing the number of ATMs and motivating people to participate in insurance programs, especially in rural areas and small cities. Opening up more commercial banks or other financial institutions, mainly in countryside areas, should also be put as one of the utmost priorities when proposing any development policies related to financial inclusion.

Appendices

Appendix 1

GNI per capita in current US$ of Vietnam from 1989 to 2021 (Unit: US$)

Year	Lower middle income - Lower bound	Lower middle income - Upper bound	Vietnam
1989	581	2,335	210
1990	611	2,465	130
1991	636	2,555	110
1992	676	2,695	130
1993	696	2,785	160
1994	726	2,895	190
1995	766	3,035	250
1996	786	3,115	300
1997	786	3,125	330
1998	761	3,030	340
1999	756	2,995	350
2000	756	2,995	380
2001	746	2,975	400
2002	736	2,935	420
2003	766	3,035	460
2004	826	3,255	530
2005	876	3,465	630
2006	906	3,595	720
2007	936	3,705	840

2008	**976**	**3,855**	**980**
2009	996	3,945	1110
2010	1,006	3,975	1360
2011	1,026	4,035	1610
2012	1,036	4,085	1970
2013	1,046	4,125	2190
2014	1,046	4,125	2380
2015	1,026	4,035	2460
2016	1,006	3,955	2570
2017	996	3,895	2700
2018	1,026	3,995	3030
2019	1,036	4,045	3280
2020	1,046	4,095	3390
2021	1,086	4,255	3560

Notes: The year that Vietnam has transitioned to a lower middle-income country are emboldened
Source: World Development Indicators database (Accessed: November 5th 2022)

Appendix 2

Top 10 provinces that have the highest average value of the financial inclusion index/Income/Relative income

City/Province	Average financial inclusion index by province			City/Province	Average households' income by province (Unit: 1,000 VND)			City/Province	Average households' relative income by province		
	2018	2016	2014		2018	2016	2014		2018	2016	2014
City - Hồ Chí Minh	2.0306	1.6554	1.4286	City - Đà Nẵng	246,809.95	204,867.46	163,799.21	Province - Tuyên Quang	1.0353	1.0310	1.0590
City - Đà Nẵng	1.9229	1.5498	1.1545	Province - Bắc Ninh	245,692.12	190,112.66	146,297.39	Province - Long An	1.0257	0.9963	0.9955
Province - Bình Dương	1.6350	1.3271	1.0968	City - Hồ Chí Minh	233,959.60	192,861.82	178,553.89	Province - Bắc Kạn	1.0203	1.0330	1.0294
City - Hà Nội	1.5410	1.2986	1.1729	City - Hà Nội	230,040.26	191,029.85	165,620.21	Province - Hoà Bình	1.0097	1.0039	0.9942
City - Hải Phòng	1.4072	1.2590	0.6524	Province - Bình Dương	219,785.90	170,202.07	160,543.81	Province - Yên Bái	1.0094	1.0323	1.0256
Province - Quảng Ninh	1.3513	1.4473	1.0769	Province - Đồng Nai	205,597.66	172,539.60	137,298.81	Province - Hà Giang	1.0089	1.0885	1.1729
Province - Bắc Ninh	1.1971	0.8790	0.6441	City - Hải Phòng	193,266.27	151,527.82	132,409.49	Province - Tây Ninh	1.0084	0.9977	1.0190
Province - Đồng Nai	1.1341	0.9454	0.7143	Province - Quảng Ninh	178,909.75	158,709.02	137,024.32	City - Đà Nẵng	1.0081	1.0103	1.0082
Province - Tây Ninh	1.0889	0.6312	0.4348	Province - Tây Ninh	173,981.36	127,064.49	107,972.72	Province - Hà Tĩnh	1.0079	1.0033	1.0220
Province - Hải Dương	1.0579	0.9581	0.6628	Province - Vĩnh Phúc	168,580.45	132,284.23	111,892.03	Province - Bình Phước	1.0078	1.0151	0.9960

Source: Author's calculations using Stata

References

Amidžic, G., Massara, M. A., & Mialou, A. (2014). Assessing countries' financial inclusion standing-A new composite index. International Monetary Fund.

Balakrishnan, M. R., Steinberg, M. C., & Syed, M. M. H. (2013). *The elusive quest for inclusive growth: Growth, poverty, and inequality in Asia*. International Monetary Fund.

Barslund, M., & Tarp, F. (2008). Formal and informal rural credit in four provinces of Vietnam. *The Journal of Development Studies*, *44*(4), 485-503.

Bauer, S. (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.

Brady, D. S., & Friedman, R. D. (1947). Savings and the income distribution. In *Studies in income and wealth* (pp. 247-265). NBER.

Bruhn, M., & Love, I. (2014). The real impact of improved access to finance: Evidence from Mexico. The Journal of Finance, 69(3), 1347-1376.

Cámara, N., & Tuesta, D. (2014). Measuring financial inclusion: A muldimensional index. *BBVA Research Paper*, (14/26).

Cámara, N., & Tuesta, D. (2017). Measuring financial inclusion: A muldimensional index. *BBVA Research Paper*, (14/26).

Demirgüç-Kunt, A., & Klapper, L. F. (2012). Measuring financial inclusion: The global findex database. World bank policy research working paper, (6025).

Demirgüç-Kunt, A., & Levine, R. (2009). Finance and inequality: Theory and evidence. *Annu. Rev. Financ. Econ.*, *1*(1), 287-318.

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution. *The World Bank*.

Frank, E. H. (2015). Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis.

Fungáčová, Z., & Weill, L. (2015). Understanding financial inclusion in China. *China Economic Review*, *34*, 196-206.

Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *The review of economic studies*, *60*(1), 35-52.

Gujarati, D. N. (2011). *Econometrics by example* (Vol. 1). New York: Palgrave Macmillan.

Jauch, S., & Watzka, S. (2016). Financial development and income inequality: a panel data approach. *Empirical Economics*, *51*(1), 291-314.

Kim, D. W., Yu, J. S., & Hassan, M. K. (2018). Financial inclusion and economic growth in OIC countries. Research in International Business and Finance, 43, 1-14

Kim, J. H. (2016). A study on the effect of financial inclusion on the relationship between income inequality and economic growth. *Emerging Markets Finance and Trade*, *52*(2), 498-512.

La, B. T., Lim, S., Cameron, M. P., Tran, T. Q., & Nguyen, M. T. (2021). Absolute income, comparison income and subjective well-being in a transitional country: Panel evidence from Vietnamese household surveys. *Economic Analysis and Policy*, *72*, 368-385.

Li, L. (2018). Financial inclusion and poverty: The role of relative income. China Economic Review, 52, 165-191.

Markus Christen & Ruskin Morgan, (2005). Keeping Up With the Joneses: Analyzing the Effect of Income Inequality on Consumer Borrowing. *Quantitative Marketing and Economics (QME)*, Springer, vol. 3(2), pages 145-173, June.

Nguyen, L. T., Hoai Nguyen, A. P., Van Passel, S., Azadi, H., & Lebailly, P. (2018). Access to preferential loans for poverty reduction and rural development: evidence from Vietnam. *Journal of Economic Issues*, *52*(1), 246-269.

Nguyen, N.T., Nguyen, H.S., Ho, C.M. and Vo, D.H., 2021. The convergence of financial inclusion across provinces in Vietnam: A novel approach. Plos one, 16(8), p.e0256524.

Park, C. Y., & Mercado Jr, R. (2018). Financial inclusion, poverty, and income inequality. *The Singapore Economic Review*, *63*(01), 185-206.

Prete, A. L. (2013). Economic literacy, inequality, and financial development. *Economics Letters*, *118*(1), 74-76.

Sahay, R., Cihak, M., N'Diaye, P., Barrajas, A., Mitra, S., Kyobe, A., ... & Yousefi, S. R. (2015). Financial inclusion: Can it meet multiple macroeconomic goals? IMF Staff Discussion Note. International Monetary Fund.

Sarma, M. (2008). Index of financial inclusion (No. 215). Working paper.

Sarma, M. (2012). Index of Financial Inclusion–A measure of financial sector inclusiveness. Centre for International Trade and Development, School of International Studies Working Paper Jawaharlal Nehru University. Delhi, India.

Sarma, M., & Pais, J. (2011). Financial inclusion and development. Journal of international development, 23(5), 613-628.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development*. International Monetary Fund.

Tran, T. Q., & Dinh, V. T. T. (2021). Provincial governance and financial inclusion: micro evidence from rural Vietnam. *International Public Management Journal*, *24*(6), 792-812.

Wang, X., & Guan, J. (2017). Financial inclusion: measurement, spatial effects and influencing factors. *Applied Economics*, *49*(18), 1751-1762.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, *6*(1), 58-82.

Williams, R. (2010). Fitting heterogeneous choice models with oglm. *The Stata Journal*, *10*(4), 540-567.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, *12*(2), 308-331.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7-20.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

World Bank (2022). The World by Income and Region. *The World Bank Database - World Development Indicators*. https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html

Zhang, Q., & Posso, A. (2019). Thinking inside the box: A closer look at financial inclusion and household income. *The Journal of Development Studies*, *55*(7), 1616-1631.

Notes

All calculations in this research paper are performed using Stata 17. Figures 1.1, 5.1, 5.2, and 5.3 are illustrated by Tableau using exported data from Stata and World Bank data.