MCom

Money Talks: Investigating the relationship between linguistic diversity and financial inclusion

A Dissertation

presented to

The **Development Finance Centre (DEFIC)**
Graduate School of Business
University of Cape Town

In partial fulfilment
of the requirements for the Degree of
Master of Commerce in Development Finance

By

Matthew Christopher, Wolf

WLFMAT002

September 2020

Supervisor: Assoc./Prof. Abdul Latif Alhassan

Contents

Plagiarism Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I certify that I have read and understand the Commerce Faculty Ethics in Research Policy (http://www.commerce.uct.ac.za/pages/ComFac-Downloads).
3. I certify that I have read the General Rules and Policies Handbook (Handbook 3) regarding Student Rules of Academic Conduct: RCS1.1 to RCS3.2 and Rules Relating to Examinations G20.1 to G22.2.
4. I certify that I have read and understand the document, "Avoiding Plagiarism: A Guide for Students".
5. I have used a recognized convention for citation and referencing. Each contribution to, and quotation in this thesis from the work(s) of other people has been attributed, and has been cited and referenced.
6. This dissertation is my own work.
7. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his own or her own work.
8. I acknowledge that copying someone else's assignment or essay or part of it is wrong and declare that this is my own work.

Signature: Signed by candidate

Matthew Christopher Wolf

List of Abbreviations

Abbreviation	Meaning
ALE	Alesina et. al Ethnolinguistic Fractionalization Index (Alesina et al., 2003)
ATM	Automated teller machines
BP Test	Breusch-Pagan Test
CIA	Central Intelligence Agency (of the United States of America)
EGIDS	Expanded Graded Intergenerational Disruption Scale
ELF	Ethnolinguistic Fractionalization
FI	Financial inclusion
FRN	Fearon Linguistic Diversity Index (Fearon, 2003)
GDP	Gross domestic product
GDPPC	Gross domestic product per capita
GNP	Gross national product
GPFI	Global Partnership for Financial Inclusion
IFI	Index of financial inclusion (Sarma, 2008)
IMF	International Monetary Fund
JB Test	Jarque-Bera Test
LD	Linguistic Diversity
LLG	Largest language group %
LPC	Languages per capita
MFII	Multidimensional financial inclusion index
NPL	Non-performing loan
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
PCA	Principal components analysis
PRC	People's Republic of China
PREG	Politically relevant ethnic groups
SUR	Seemingly unrelated regression
USA	United States of America
UNESCO	United Nations Education, Scientific, and Cultural Organization
UK	United Kingdom
UNSD	United Nations Statistics Division

List of Tables

List of Equations

Acknowledgements

I am grateful to my partner and my family for their unwavering support, inexhaustible patience, and unconditional love during my studies. I could not have done this without them. I am particularly grateful to for the supervision, the guidance, and – above all – the patience of Assoc./Prof. Abdul Latif Alhassan during the research process. Thank you.

Abstract

The differences in languages spoken within a population can be thought of as transaction costs on the economic activities of that population. This perspective has motivated a host of academic literature analyzing how countries' linguistic profiles relate to different socio-economic variables. Among these studies, financial inclusion is rarely one of the variables of interest. Language and financial inclusion are sometimes analyzed together in more granular studies of a single country, or even of individuals, but never in cross-sectional, country-level analyses. However economic growth, which is generally considered to be positively related to financial inclusion, has frequently been studied in relation to language, with mixed results. Earlier researchers of the question identified negative relationships between economic growth and linguistic diversity, in what became known as the "Fishman-Pool Hypothesis". Later researchers determined that such a relationship did not exist, or that, in certain contexts, linguistic diversity and economic growth could even be positively related. This study departs from the intuition that financial inclusion's relationship to linguistic diversity may parallel that of economic growth – a relationship that seems intuitively negative but is more ambiguous after analysis. To overcome the broad interpretability of the concepts of interest, this study constructed two dependent variables representing financial inclusion, and four independent variables representing linguistic diversity with cross-sectional data for a sample of 86 countries.

The models were estimated by accounting for multicollinearity of the regressors, as well as heteroskedasticity and non-normality in the error terms using the Seemingly Unrelated Regressions models and ordinary least squares estimation techniques. The results indicate that linguistic diversity indicators were all nearly zero and highly insignificant despite the strong specification of the models. This suggests that linguistic diversity has no significant relationship – positive or negative – to financial inclusion at a country level. This result was consistent across all the possible combinations of the operationalized variables for both concepts.

Chapter 1: Introduction

1.1 Background of the Study

Language is a crucial element of all human interactions and economic transactions. For two parties to engage in a transaction, they need to be able to communicate in a mutually understood language. The assumption of a common language between parties is often justified, as parties who speak different languages may still transact through translators, intermediaries, or in a *lingua franca*, a language used between people whose main languages are different. But even then, parties may remain uncertain about intended and perceived meaning of their communication. As a result, linguistic differences between individuals or groups could be considered a kind of transaction cost.

Researchers have confirmed this idea of linguistic difference as transaction cost. It has been shown that language considerations influence choices made in bilateral trade and international investment decisions - such as foreign firm acquisitions and foreign securities investments - in a manner consistent with transaction cost theory (Hutchinson, 2005; Isphording & Otten, 2013; Vidal-Suárez & López-Duarte, 2013). This makes sense: these economic activities take place in an international – and so, possibly multilingual – context. Thus, language and language differences can and do influence international economic decisions.

Linguist Joseph Greenberg also suggested that language may also affect domestic economics and national development as well: in a paper on measuring linguistic diversity, he tangentially professed an "expectation, subject to significant qualifications," that linguistically diverse polities would likely not yet be economically developed or politically organized - otherwise a single lingua franca would have already been adopted in these polities to facilitate more effective communication (Greenberg, 1956).

Researchers subsequently interrogated this "expectation", trying to understand the relationship, if one exists, between linguistic diversity and other social, political, and economic outcomes. Initial research into the connection between these variables supported Greenberg's "expectation". What became known as the "Fishman-Pool Hypothesis", named for two of the earlier researchers into this question, corroborated that a negative relationship did indeed seem to exist between measures of linguistic diversity and economic growth or development. This implied that linguistically heterogeneous countries were either less developed, or developed more slowly than their linguistically homogeneous counterparts (Fishman, 1964; Nettle et al., 2007; Pool, 1972). This seems to suggest that language uniformity could act as a catalyst to economic activity within a country, while language diversity could act an inhibitor. But researchers have challenged this interpretation in different ways. They have at times concluded that no relationship exists between these concepts (J. Arcand & Grin, 2013; J. L. Arcand, 1996; Nettle, 2000), or that, when analysed with proper controls and different scopes, there may be a robust case for localized economic benefits to higher linguistic diversity (J. Arcand & Grin, 2013; J. L. Arcand, 1996; Desmet et al., 2016).

Unlike linguistic diversity, financial inclusion's relationship to economic development is not ambiguous in academic literature. In general terms, research has shown that economic development is positively related to financial inclusion. (Dabla-Norris et al., 2015a). In addition, greater financial inclusion is shown to be positively associated with many measures of socioeconomic development, such as higher income, lower inequality, and higher literacy rates (Chikalipah, 2017; Sarma & Pais, 2010b).

Despite the ambiguity of the relationship between linguistic diversity and economic development and the clarity of the relationship between economic development and financial inclusion, very few researchers have thought to analyse the relationship between linguistic diversity and financial inclusion. The reason for this may lie in the difficulty of instrumentalizing both concepts. "Financial inclusion" and "linguistic diversity" as concepts are open to interpretation, and can be measured and modelled in different ways (Cámara & Tuesta, 2017; Desmet et al., 2016; Greenberg, 1956; Laitin & Ramachandran, 2014; Pool, 1972; Sarma, 2008). This makes it difficult to determine if any observed relationship between these two concepts is real, or just the spurious product of the choices made in operationalizing variables to represent them. The studies that include both language and financial inclusion in their analyses tend to use individual consumers or providers of financial services as the unit of analysis, often in multilingual countries or communities, as in South Africa (Wentzel, Diatha, & Yadavalli, 2016), or amongst recent immigrants to the United States (United States Government Accountability Office, 2010). In these cases, consumers' native language is often one among many different independent variables analysed, rather than the sole focus of the study. There seem to be no studies that analyse the specific relationship of language to financial inclusion at a country level.

This raises the question: Does a relationship exist between the linguistic diversity of a country and the level of financial inclusion in that country? Answering this question will be the primary aim of this study.

1.2 Research Problem

As described above, literature on the connection between linguistic diversity and economic development at a country level is divided – some research supports the idea that the two concepts are negatively related (Desmet et al., 2016; Fishman, 1966; Nettle, 2000; Nettle et al., 2007; Pool, 1972), while other research supports the idea that they are positively related under certain circumstances (J. L. Arcand, 1996), and still others hold that no relationship exists at all (Laitin & Ramachandran, 2014; Lian & Oneal, 1997). On the other hand, it is generally accepted that financial inclusion and economic development are positively related (Chikalipah, 2017; Dabla-Norris et al., 2015a; Sarma & Pais, 2010b). However, no cross-country research which specifically investigates the relationship between linguistic diversity and financial inclusion exists. This gap in research represents a lost opportunity for academics, development practitioners, policy makers, and financial service providers to better understand how language may affect efforts to increase financial inclusion and how it affects economic development more broadly. Research on the nature of this relationship could open new avenues for collaboration in linguistics and development finance, as well as contribute nuance to the broader conversation about language policy and economic development.

This study investigates the relationship between linguistic diversity and financial inclusion. It will seek to answer these questions: is there a relationship between a country's levels of linguistic diversity and financial inclusion? If so, what is the nature of this relationship?

1.3 Research Objectives and Hypothesis

This study has one objective:

To examine the relationship between financial inclusion and linguistic diversity.

The null hypothesis of the study is that no significant relationship will be found between measures of financial inclusion and linguistic development at a country level.

As shall be discussed in more detail, financial inclusion and linguistic diversity can be interpreted and operationalized in many ways. For this reason, this study will test for relationships between several operationalizations of both concepts – four variables representing linguistic diversity, and two representing financial inclusion. By testing the hypothesis with different combinations of these operationalized variables, this study aims to account for the broad interpretability of the concepts of interest, and test whether any relationship (or lack thereof) between them persists across different operationalized variables. Thus, this hypothesis will be tested eight times - four times two combinations of variables.

1.4 Justification of the study

The justification of this study is that it will begin to fill the gap in existing literature on the relationship between linguistic diversity and financial inclusion. This piece of research will be a small step towards filling this gap that connects language to concepts in development finance, hopefully motivating further, more nuanced research into these concepts' relationship.

The study also has practical applications for policy makers, financial service providers, development practitioners, and other social scientists. Its question is particularly important in linguistically diverse, less-developed countries, especially those in which the official or institutional language is not universally used, as in many formerly colonized countries. Sub-Saharan Africa, for example, exhibits some of the lowest measures of formal financial inclusion worldwide (Demirgüç-Kunt et al., 2018) and is one of the most linguistically diverse regions on Earth (UNESCO, 2006) A better understanding of the relationship - or lack thereof - between linguistic diversity and financial inclusion will help financial service providers and policy makers to adapt their businesses and policies to achieve greater financial inclusion, particularly in linguistically diverse regions. This may be particularly relevant in places experiencing high degrees of human migration, where the mix of languages spoken within a population is shifting rapidly, and local governments need to act quickly to integrate new immigrants.

Finally, this research is important because it may inform our understanding of what appears to be a trade-off between linguistic diversity and aspects of economic development. Global linguistic diversity is falling worldwide, particularly in less developed parts of the world (Loh

& Harmon, 2010). As will be discussed in the literature review below, much research suggests that linguistic diversity is negatively correlated with measures of economic development and growth (Fishman, 1964; Nettle et al., 2007; Pool, 1972). Thus, a less linguistically diverse world may be a natural consequence of global economic development, or even a desirable one in the opinions of some. However, some literature suggests the opposite – that linguistic diversity is not related, or is even positively related to some desirable economic outcomes (Laitin & Ramachandran, 2014; Nettle, 2000; Ottaviano & Peri, 2006). Other literature points out that linguistic diversity has social and economic value as well (Grin, 2007; Skutnabb-Kangas et al., 1999). So, is there really a trade-off between linguistic diversity and aspects of economic development? If so, how do we balance the value of that development against the value of the diversity? These questions are beyond the scope of this study. However, it is hoped that this study will shed some small light on these larger issues, and thereby allow countries to both increase financial inclusion and preserve cultural and linguistic heritages.

1.5 Organization of the Study

This study is organized into the following four chapters, beginning after this first chapter:

The second chapter is a literature review, in which the concepts of linguistic diversity and financial inclusion are defined, and existing theoretical frameworks for understanding both are examined. The chapter also considers several existing conceptualizations of both ideas, to illustrate the difficulty and variety with which they are measured in the academic literature. Finally, it also examines several past studies on the relationship between economic development and linguistic diversity, between economic development and financial inclusion, and, broadly speaking, between language and financial access.

The third chapter outlines the methodology of the study. This includes the sources, period, and samples of the country-level data used, the analytical approach and steps taken to answer the research question, and an explanation of the different operationalizations of linguistic diversity and financial inclusion, as well as the selected control variables.

The fourth chapter explores the findings of the study. The fifth chapter summarizes the study, its limitations, the possible recommendations to draw from its results, and the apparent avenues for further research on its subject.

Chapter 2: Literature Review

2.1 Introduction

This research unites two traditionally separate areas of study. As mentioned, the overlap in the literature on linguistic diversity and financial exclusion is sparse. However, there is a great deal of literature examining the connections between financial exclusion and economic development, and a smaller body of literature connecting linguistic diversity and economic development. The methodology of this proposed research draws greatly on the work in these two bodies of literature. This literature review covers:

- definitions of financial inclusion and linguistic diversity and how this study tries to overcome these concepts' "broadness",
- how linguistic diversity and financial inclusion each relate to economic development in the literature,
- the ways that researchers conceptualize and operationalize variables to measure these imprecise concepts,
- And a selection of empirical works that help inform our understanding of how financial inclusion and linguistic diversity appear to be related.

2.2 Definition of concepts and terms

"Financial inclusion" and "linguistic diversity" are broad concepts that can be defined and conceptualized in many ways. This section discusses the difficulties of trying to contain either concept in a single variable. In response to these difficulties, this study's methodological approach will use different variables to operationalize both concepts. A short discussion of how both these concepts are frequently defined is instructive and helpful to understanding this approach.

One of the earliest definitions of "financial inclusion" in academic literature refers to its opposite, financial exclusion. In 1995, Leyshon and Thrift wrote about how the decisions of banks in the UK and USA led to financial exclusion, which they describe as "those processes that prevent poor and disadvantaged social groups from gaining access to the financial system" (1995). The inverse – processes that *enable* poor and disadvantaged social groups to gain access to the financial system – can therefore be taken as their implicit definition of financial *inclusion*.

Other researchers have also described financial inclusion as processes (Conroy, 2005; Mohan, 2006). This process-orientation is useful for thinking about what can influence, either positively or negatively, the accessibility and uptake of formal financial services. Financial inclusion can also be defined as a static, demographic characteristic of individuals or households, which can then be measured in aggregate for populations. The famous World Bank Global Findex data does just this – two of its creators, Asli Demirgüç-Kunt and Leora Klapper, defined financial inclusion simply as the "use of formal financial services" (2013). Later in 2018, in the report accompanying the release of the 2017 Findex data, the same authors and their collaborators expanded this definition to "access to and use of formal financial services" (Demirgüç-Kunt et al., 2018). Using this measurement, financial inclusion has increased worldwide over the past seven years, as evidenced by an increase from 51% of adults worldwide with an account in 2011 to 69% in 2017 (Demirgüç-Kunt et al., 2018). They also redefined "account ownership" to include having either a formal bank account *or* a mobile money account. The inclusion of mobile money accounts in the revised definition stems from the fact that there has been substantial growth in the use of mobile money as a means of accessing financial services, particularly in Sub Saharan Africa, and in some other countries like Bangladesh (Demirguc-Kunt et al., 2018). In these regions where digital financial services and similar innovations are taking root, measuring financial inclusion by simple possession of a bank account may no longer indicate the reality on the ground (Demirguc-Kunt et al., 2018). These varied and changing definitions of financial inclusion show how it is a concept interpretable in many ways, and in a constant state of flux as the financial services industry changes with time.

Similarly, "linguistic diversity" can have different definitions. All aim to quantify the number of languages and degree of difference in the languages spoken by a population of interest. The most common definition is "language richness" - the number of languages spoken in a population or geographic area – likely because it is simplest to measure (Loh & Harmon, 2010). However, this simple approach ignores practical considerations, such as the possibility of multilingualism, or mutual intelligibility between languages (Greenberg, 1956). To attempt to account for these, one definition counts the number of language families present in a population's spoken languages, also known as phylogenetic diversity. Another quantifies the structural, grammatical differences between the population's languages, also known as structural diversity (Loh & Harmon, 2010). Still others measure how uniform or concentrated the distribution of languages is among members of a population – called "language evenness" (Loh & Harmon, 2010). Early studies of linguistic diversity tended to focus exclusively on simple Herfindahl indices of language richness – see section 2.5 for more on these (Greenberg, 1956; Herfindahl, 1950). To better capture linguistic differences, and not just numbers of languages, some recent research focused on economic development and linguistic diversity has used measures of phylogenetic diversity (Desmet et al., 2009; Hammarstrom, 2016). However, there is no universal definition of what linguistic diversity means – most definitions tend to be "fit-to-purpose", whether that purpose is political science research, or endangered language preservation. Nearly all the different definitions of linguistic diversity suffer from a serious lack of data (J. L. Arcand, 1996). Country-level language data is not always available, and often is only collected as frequently as a national census occurs. When this country-level data is collected, it often only includes respondents' home languages, or whether they speak a national or official language, rather than detailed information on what languages they speak, and to what degree they are fluent. The most frequently cited source when measuring linguistic diversity is Ethnologue, a non-profit that studies and documents languages to promote literacy (Simons & Fennig, 2018). Ethnologue, in addition to other indices compiled by past researchers, will be the primary source of linguistic diversity data for this study.

This study adopts the following working definitions:

- Financial Inclusion: access to and usage of formal financial services, including mobile money services, whether they are linked to a formal financial institution.
- Linguistic Diversity: the quantity of, and degree of difference between, the languages spoken as native, home languages by the members of a population of interest.

These definitions are intentionally broad: in this study, different conceptualizations of both concepts are used to reflect different aspects of both financial inclusion and linguistic diversity, and to observe how this in turn affects the outcomes of our analyses. The definitions are intended to encompass these different conceptualizations.

2.3 Theoretical perspectives

2.3.1 Linguistic diversity and economic development:

The literature about the relationship between linguistic diversity and economic development centres on an important question which mirrors this study's own research question: is there a relationship between these two concepts? If so, what is its nature? Joseph Greenberg began this line of inquiry with his seminal work, *The Measurement of Linguistic Diversity* (Greenberg, 1956). In it, he posited that areas with less linguistic diversity would be more economically productive and politically united – hence his aforementioned "expectation" that his indices measuring linguistic diversity would correlate negatively with countries' levels of economic development and growth (Greenberg, 1956). As he saw it, increased development would create a greater need for a common language as well as economic incentives for individuals to learn this common language, which would erode linguistic diversity over time (Greenberg, 1956).

Subsequent tests of Greenberg's assertion seem to prove him right. Different studies' comparisons of "linguistically homogeneous and linguistically heterogeneous" countries across different measures of development – such as GNP, government revenues, and education enrolment –indicated that greater economic and social development are positively correlated with lower levels linguistic diversity (Fishman, 1966; Nettle, 2000; Pool, 1972). Studies comparing countries' linguistic diversity to their per-capita GDP (GDPPC) demonstrated that there were poor countries with low linguistic diversity, but there were almost no wealthy countries with high linguistic diversity (Pool, 1972). Thus Greenberg's "expectation" became the "Fishman-Pool hypothesis" and asserted that high linguistic diversity and high economic development were inversely related.

However, this conclusion is contested by some researchers. Newer studies have substituted economic growth and political stability for GDP and GDPPC and found no significant relationship between these variables and linguistic diversity (Lian & Oneal, 1997). Other studies tested the Fishman-Pool hypothesis in different country income bands and controlled for GDPPC levels to see whether linguistic diversity still seemed to impact other, social measures of quality of life. It was found that the negative relationship found by Fishman and Pool was still valid, though less explanatory than previously believed (Nettle, 2000).

Thus, studies testing the Fishman-Pool Hypothesis have been mixed in their conclusions about the true nature of the relationship between linguistic diversity and economic development. For instance, Daniel Posner had strong criticisms of measuring ethnolinguistic diversity using "ethnolinguistic fractionalization" – an approach to operationalizing linguistic and ethnic diversity - and so devised an index that grouped language groups based on their political alignments in their respective countries (called PREG, for Politically Relevant Ethnic Groups). However, his new index exhibited a similar negative relationship with measures of economic development (Posner, 2004). On the other hand, economist Jean-Louis Arcand applied an Instrumental Variable approach, of the kind pioneered by Acemoglu, Johnson, and Robinson, using country-level English language proficiency as the instrument between economic development (as the dependent variable) and linguistic diversity (as the independent variable), and showed that there appeared to be a slight positive relationship between use of local languages and economic development in formerly colonized countries (Acemoglu et al., 2001; J. Arcand & Grin, 2013). Kouame offers a more theoretical basis for why this may be – noting that forced adoption of colonial languages in African business contexts is economically exclusionary for speakers of local languages, and that embracing these local languages can be winning strategy for catalysing local growth (Kouame, 2014). Interestingly, he then applies this theoretical framework to microfinance in Côte d'Ivoire, in a case examined further in section 2.7.4.

As illustrated above, the relationship between economic development and linguistic diversity is still debated in academic literature. Researchers have used a range of innovative methodological approaches to approximate linguistic diversity, and these approaches offer many opportunities for re-application in similar studies. Despite the division of opinion, the majority of the literature appears to confirm Greenberg's initial expectation that there is a negative relationship between linguistic diversity and economic development *on a country level*, though no researcher goes so far as to assert any causality in this relationship (Fishman, 1964; Nettle, 2000; Pool, 1972; Posner, 2004). At the same time, there is a smaller body of empirical work that suggests that local linguistic diversity may actually benefit growth on a smaller scale, at the level of communities and firms (J. Arcand & Grin, 2013; Carletti et al., 2012; Kouame, 2014).

2.3.2 Financial inclusion and economic development

The literature on the relationship between economic development and financial exclusion is far larger and far less contentious than that on the relationship between economic development and linguistic diversity. It is generally accepted that financial inclusion benefits individuals, by more easily facilitating investment in productive activities like education and investment (Demirgüç-Kunt & Klapper, 2013), enabling asset building, wealth creation, and consumption smoothing, as well as reducing risks and transaction costs (Allen, Demirgüç-Kunt, et al., 2012; Sarma & Pais, 2010b). Studies also show that policy action designed to enable greater financial access to and usage of financial services has unambiguously positive effects on economic growth and development (Dabla-Norris et al., 2015a, 2015b). Any factor that had a positive effect on financial inclusion could be expected to have an indirect positive effect on economic development, assuming it had no other, offsetting, direct or indirect effects through other means.

2.4 Measuring linguistic diversity

Literature concerning linguistic diversity has its origins in anthropologists' and sociologists' efforts to differentiate between cultures or ethnic groups. Quantifying these differences has long been a challenge these researchers have faced. Greenberg's *The Measurement of Linguistic Diversity* laid out several other methods for "developing quantitative measures of diversity" that rely on detailed data on population language data which, as mentioned, is often not available (J. L. Arcand, 1996; Greenberg, 1956). At their core, these methods measure the probability of two randomly selected members of a population being able to communicate in a common language. The quantitative methodology employed is identical to that of the more commonly known Herfindahl indices used in economics to measure the market power of companies in an industry (Herfindahl, 1950). These "Greenberg indices", as they are now called in linguistic contexts, can be modified to account for multilingualism in studied populations, as well as degrees of resemblance and mutual intelligibility (language "distance") between languages considered (Greenberg, 1956). They cannot, however, satisfactorily account for the varying levels of proficiency of a population's multilingual individuals. The main impediment facing those that would construct a robust Greenberg index is the lack of population-representative data about multilingualism – or indeed, about any languages spoken at all (J. L. Arcand, 1996). Census data sometimes includes data about individuals' home or native language, but rarely any data about the other languages they may speak, particularly in developing countries (Fishman, 1964; Nettle, 2000). As a result, simple monolingual Greenberg indices tend to overestimate the amount of linguistic diversity in a population.

The lack of representative language data for populations and the difficulty of creating such data means that researchers rarely construct their own Greenberg indices. Rather, they often use proxies for linguistic diversity. One of the simplest measures of linguistic diversity, which was employed by several early studies on the effects of language on development, is the size of the largest native-language community in a country as a percentage of the country's total population (Pool, 1972). Using this measure, a lower value indicates "higher" linguistic diversity. However, such a measure does not conceptualize "linguistic diversity" realistically, since it also does not account for multilingualism in its population, nor does it account for the possibility that different languages in a country might be similar enough to be mutually intelligible, as some more advanced Greenberg indices do (Greenberg, 1956). It's simplicity also belies two analytical drawbacks: it doesn't account for the geographical size of countries, and it tends to overstate linguistic diversity for formerly colonized states, whose modern boundaries don't reflect traditional ethno-cultural boundaries (Nettle, 2000). Nettle used another simple measure to overcome the first drawback – the number of living languages in a country divided by the country's population in millions (Nettle, 2000). It is important to note that both Greenberg's indices and Nettle's languages-per-capita both measure language richness as described in Section 2.2 – but not phylogenetic or structural diversity (Loh & Harmon, 2010).

Despite the difficulty of building Greenberg indices, a handful of dedicated researchers have created broad, country-level datasets using Greenberg's methodologies or variations thereof, often using the aforementioned "ethno-linguistic fractionalization" approach, more generally called ELFs. (Alesina et al., 2003; Desmet et al., 2009; Fearon, 2003). Several such datasets have been created since the early 2000s. The most important include Alesina's indices for ethnic, religious, and linguistic diversity, Fearson's index for cultural diversity based on intra-linguistic "distance", and Desmet et al's indices based on phylogenetic language trees (Alesina et al., 2003; Desmet et al., 2016; Fearon, 2003). The former two of these are repurposed in this study's methodology.

2.5 Measuring financial exclusion

The wealth of data on financial inclusion and the vibrant history of academic work on the subject has led to a proliferation of measurements, indices, and methodologies for assessing financial inclusion. The simplest is simply counting individuals with a formal bank account (or, more recently, a mobile money account) as "financially included", and checking the proportion of such individuals in a population (Demirguc-Kunt et al., 2018; Demirgüç-Kunt & Klapper, 2013). The Global Partnership for Financial Inclusion measures financial inclusion using fifteen different indicators to capture three metrics: access to financial services, usage of financial services, and the quality of financial products and service delivery. (Global Partnership for Financial Inclusion, 2012).

This variety of measurements is due to the diversity of financial systems, products, and services around the world through which people can become “included”. For example, while Sub Saharan Africa has some of the lowest rates of formal banking penetration in the world (33% of adults with a formal account vs 67% worldwide), they also have some of the highest rates of mobile money usage in the world (21% of adults with a mobile money account vs 4% worldwide) (Demirgüç-Kunt et al., 2018).

Despite the consensus amongst researchers on the relationship between economic development and financial inclusion, financial inclusion has one thing in common with linguistic diversity: it is challenging to conceptualize and operationalize in a single, comprehensive variable. However, unlike linguistic diversity, there is a plethora of country-level data collected and publicly available which researchers can use to try to quantify financial inclusion, including the World Bank Group’s Global Findex data, Enterprise Surveys, Global Payment Systems Surveys, or the IMF Financial Access Survey data (Demirgüç-Kunt et al., 2018; The International Monetary Fund, 2018; The World Bank Group, 2016a, 2016b). This abundance of data has allowed researchers to quantify and measure financial inclusion in a number of simple ways, including possession of a formal bank account, use of savings products, and prevalence of borrowing from banks (Demirgüç-Kunt & Klapper, 2013). Additional details about frequency of and reasons for using financial products, modes of access, and barriers to use allow researchers to build customized indices of financial inclusion, such as those proposed by the Global Partnership for Financial Inclusion (Global Partnership for Financial Inclusion, 2012). Other researchers, often focused on country-specific or community-specific populations, collect their own data to assess financial inclusion (United States Government Accountability Office, 2010; Wentzel, Diatha, & Yadavalli, 2016).

Two financial inclusion indices worth noting for their flexibility and comprehensiveness are the Index of Financial Inclusion (IFI), developed by Sarma, and the Multidimensional Financial Inclusion Index (MFII) built by Cámara and Tuesta (Cámara & Tuesta, 2017; Sarma, 2008)

Sarma developed the IFI in reaction to a perceived lack of a comprehensive measurement of financial inclusion that could be used across country contexts and in comparison to other development-related variables (Sarma, 2008). The IFI is flexible, in that it can include various components judged relevant to financial inclusion. Sarma herself includes three: the proportion of a country's population with a bank account, which approximates banking system penetration, the number of ATMs or bank branches per 1,000 people, which approximates banking system access, and the volume of credits and deposits as a proportion of national GDP to approximate banking system usage (Sarma, 2008). The IFI differs from typical development indices in that it is a Euclidean distance from the "ideal" – the best score possible in the index. Shortly after developing the IFI, Sarma used it to reconfirm the relationship between financial inclusion and development in several papers. She determined that at a country level, the IFI moved in tandem with many – though not all – generally accepted measures of economic development and human welfare, including income, inequality, adult literacy, and many infrastructure related variables (Sarma & Pais, 2010b, 2010a).

While the IFI offers a useful approach to measuring financial inclusion, it has some weaknesses. Since its components are still chosen by the researcher, it can only operationalize aspects of financial inclusion for which there exists robust, cross-sectional data for many countries at given point in time. However, since Sarma first proposed the IFI, many such data resources have come to exist, making her index more useful than ever.

Cámara and Tuesta sought to create a similar index of financial inclusion, but with a different statistical basis and with stronger incorporation of "demand-side" indicators of financial service uptake (Cámara & Tuesta, 2017). They criticize Sarma's approach as a parametric index – one in which the weights on index components are assigned arbitrarily by the index builder - and as reliant on only supply-side data. The IFI is sensitive to these subjective choices which "assume" the model of the data analysed and do not fully reflect demand-side factors that influence financial inclusion. Cámara and Tuesta proposed a non-parametric approach that estimates the weights that should be applied to the index components based on the raw data's latent structure, in a process called principal component analysis (PCA) (Cámara & Tuesta, 2017). They do this twice – once to estimate weights for three sub-indices, representing usage of financial services, access to financial services, and barriers to financial services, and a second time to discover the weights for each of these sub-indices to create the final MFII. This approach helps reduce researcher bias in the outcome of index construction. In addition, Cámara and Tuesta incorporate demand-side measurements of financial inclusion (those under the 'barriers' sub-index) from Findex data, not previously available when Sarma created the IFI (Cámara & Tuesta, 2017; Demirgüç-Kunt et al., 2018). This methodological approach is useful in that it also allows one to determine how important each sub-index and component of the overall MFII is to financial inclusion overall, since PCA gives us useful information about which factors generate have the most explanatory power in the data (Cámara & Tuesta, 2017).

2.6 Empirical literature

There are few studies focused specifically on the relationship between language or linguistic diversity and financial inclusion at a country level. There are some single-country studies that include language as either an independent or control variable in their analyses of financial inclusion. There are also studies that analyse financial inclusion's relationship to variables that could act as instruments or proxies for linguistic differences, such as consumers' ethnicity, their literacy, or their education levels. Still others analyse the relationship between these two concepts, but among immigrants to a country rather than the country‘s full population. Finally, some case studies of organization's efforts to reach financially excluded consumers in linguistically diverse communities offer interesting insight into how these two variables might be related. This section will consider these different groups of studies in turn.

2.6.1 Financial Inclusion Studies with Language or Linguistic Diversity as a Variable

Many studies of financial access or inclusion use language as either an independent variable or as a control variable. Two useful examples come from South Africa, where Wentzel and his collaborators showed that, for financially excluded individuals at the "bottom of the pyramid", an individual's home language was a significant determinant of their likelihood to be financially included (Wentzel, Diatha, & Yadavalli, 2016). They also analysed these consumers' preferences for accessing financial services, and found that they would prefer to access them in supermarkets, largely because of the perception that there they could receive service in their home languages (Wentzel et al., 2013). Wentzel et al. expressed surprise at home language‘s significance in the first study as a determinant of financial inclusion and suggested that future researchers should investigate the variable‘s role in South African financial inclusion further.

Linguistic diversity relates to financial inclusion at a more local geographic level in some contexts: one study of financial inclusion in Mexican municipalities used "ethno-linguistic diversity" as a control variable and measured it as the percentage of each municipality that speaks an indigenous language instead of Spanish as their native language. They determined that municipalities with higher proportions of indigenous language speakers did indeed have lower levels of financial inclusion and higher levels of income inequality, significant in all their variations of their methodology to 1% - though they refrain from attributing this to the linguistic diversity of these municipalities (García-Herrero & Martínez Turégano, 2015; Salazar-Cantú et al., 2015).

Language impacts financial access via digital financial services as well. A qualitative, focus-group study of rural mobile money users in Kenya found that one of the important challenges that users faced was lack of service in their languages. Many were fearful of using mobile money, as they knew that if there were problems they would only be attended by support staff that spoke English or Kiswahili – languages in which many of the participants felt unable to express themselves (Otieno et al., 2016). Language barriers are a challenge to broader adoption of mobile banking services, as in Ethiopia where most mobile technology is not adapted to local languages (Asfaw, 2015). It is also a barrier to deeper, more engaged use of mobile banking services, as in Tanzania, where many use mobile money transfers for casual transactions, but nothing more, due to a lack of Kiswahili language features (Mramba et al., 2014; Rumanyika, 2015). That these results appear in many single-country studies would suggest that they are somewhat generalizable across countries – however, there are counterexamples. For example, a qualitative study of Fijians' perceptions of mobile banking services showed that very few had language-related concerns about using the services (Finau et al., 2016). On the other hand, participants in this study were already bank customers who had not activated mobile banking, and 94% of them were literate. Work by the Pacific Financial Inclusion Program and other researchers indicates that language, and more generally, lack of colonial language literacy, are indeed obstacles to financial inclusion for many Pacific Islanders (Eves & Titus, 2017; Lee & Chang, 2019).

2.6.2 Financial Inclusion Studies with Potential Instruments of Language or Linguistic Diversity as Variables

The studies mentioned in the previous section are among the few that formally included language as a variable (in quantitative studies) or as a theme or area of investigation (in qualitative studies). Many other empirical studies of the determinants of financial inclusion seem to ignore language as a variable altogether in their analysis. However, there are other variables that are closely tied to, or even proxies for, linguistic differences or linguistic diversity. These variables include education levels, literacy (and/or financial literacy), gender, and ethnic diversity. We know that language is involved in these studies because the researchers, despite not having included language as a formal variable of analysis, make language-related insights or recommendations in their studies' conclusions.

The intimate relationship between language and education can best be summed up with two quotes from a UNESCO report on adapting education to local languages: "To be taught in a language other than one's own has a negative effect on learning," and conversely, "Using the home language as the language of instruction has a positive impact across the board" (UNESCO, 2016). Education and literacy are frequently studied determinants of an individual's probability to be financially included. At the same time, "using the home language as the language of instruction" in linguistically diverse countries is not always possible– both for practical reasons, such as the costliness of producing multilingual materials (Brock-Utne, 2001) or difficulty of finding multilingual teachers (UNESCO, 2016), and for historical reasons, such as colonialism, short-sighted, donor-led "literacy for all" programs, and post-colonial nationalist unification efforts (Brock-Utne, 2001; Coyne, 2015; Kosonen, 2017; Kwok, 2019). It follows then, that linguistic diversity may negatively affect financial inclusion indirectly, through its effect on education and literacy – which are frequently among the most important determinants of financial inclusion (Demirgüç-Kunt & Klapper, 2013; Evans & Adeoye, 2016; Honohan & King, 2009; Zins & Weill, 2016a). There are many studies of financial inclusion that include education or literacy variable and exclude language variables, and in which the authors reach conclusions or make recommendations specifically related to language. From these studies, we can glimpse some cases in which language is judged to be part of the cause for, or solution to, low levels of education and, in turn, financial inclusion.

Many studies determine that low education or illiteracy in the language used by a country's formal financial services providers creates an awareness barrier, leading to involuntary financial exclusion. For instance, a study of Niger specifically and the West African Economic and Monetary Union more generally determined that higher education levels are significantly correlated with higher likelihood of being formally financially included (Chaibou, 2019a). The author recommended that to overcome this barrier, banks and the government should accompany the launch of any financial inclusion initiatives with communications campaigns in local languages to better reach target consumers who would otherwise remain ignorant of available financial services (Chaibou, 2019a). However, evidence in India shows that even if this initial awareness barrier can be overcome for the financially excluded who are less-educated or illiterate, effort must be made to adapt further financial education and access points

– such as ATMs and bank marketing materials – to local languages. (Bihari, 2011; Singh et al., 2014). In the same study, the authors compare several different business models and technologies on their merits for boosting Indian financial inclusion – local language compatibility (or lack thereof) is consistently cited as an advantage (or disadvantage) of each of the different options presented (Singh et al., 2014). This "awareness barrier" can be thought of as involuntary financial exclusion through language.

Low education and literacy levels may also exacerbate voluntary financial exclusion - in Mozambique, a study determined that individuals with less education or who were illiterate in Portuguese tended to have very low trust in formal financial services, and opt out of them even when they were accessible (Jossefa, 2011). This was similarly true in Angola (Portuguese), Nigeria (English), and Côte d'Ivoire (French) (Amaeshi, 2011; Kouame, 2014; Vicente, 2011). This tendency to define "literacy" in terms of colonial languages is not uncommon – these languages are very frequently the languages of instruction in formerly colonized countries (Brock-Utne, 2001; Coyne, 2015) and often also the languages of financial services. Studies like these point to illiteracy as a key cause of financial exclusion, but they often define illiteracy terms of the one colonial language, and no other local languages. In a sense, this obstacle could be reframed as the result of colonially imposed monolingualism in a linguistically diverse context.

Other variables besides education and literacy can also act as proxies or instruments of language or linguistic diversity, such as ethnicity. In some countries, such as Mexico, certain ethnicities correspond to non-majority language groups. The previously mentioned study of Mexican municipalities is an example – it used the prevalence of indigenous languages in municipalities as a proxy for ethno-linguistic diversity (Salazar-Cantú et al., 2015). Another study in Mexico makes this even more explicit: it analyzes the determinants of financial exclusion among populations of people of indigenous "Maya" ethnicities in the Yucatan peninsula (Escalante & Batun, 2018; Salazar-Cantú et al., 2015). The researchers found that speakers of the Maya languages were significantly less financially literate, and significantly more likely to be financially excluded (Escalante & Batun, 2018). They, and other researchers of financial exclusion in Mexico, also noted that financially excluded people of indigenous ethnicities also have cultural and psychological barriers to becoming financially included – which are augmented if that ethnicity speaks a minority language (Pleite et al., 2016; Salazar-Cantú et al., 2015). The same linguistic, cultural, and psychological obstacles are a factor in efforts to financially include indigenous, minority language-speaking ethnicities in Australia (McDonnell, 2003), Bangladesh and Pakistan (Kempson & Whyley, 1999).

This type of language barrier can manifest through gender as well as ethnicity. A study of women's barriers to financial inclusion in Papua New Guinea – one of the world's most linguistically diverse countries - illustrates this: researchers concluded that the primary obstacle to poor women accessing financial services was their poor English skills, due to English's *de jure* status as the language of formal financial services in the country (Eves & Titus, 2017).

2.6.3 Studies of Language and Financial Inclusion of Immigrant Populations

There is one group of people for whom it is established that language is indeed related to – and often a barrier to - financial inclusion: immigrants. Studies in the US and the UK frequently highlight the importance of language in enabling financial inclusion. One study by the United States Social Security Administration determined that English language skills were an important determinant of an immigrant's likelihood of owning a formal bank account or any other financial asset in the USA (Barcellos et al., 2012). Notably, immigrants with worse English language skills exhibited lower confidence in the trustworthiness of formal financial institutions. These relationships were unaffected by individuals' sex, region of origin, age, or any other demographic characteristics – suggesting that the language barrier to financial inclusion was real and pervasive (Barcellos et al., 2012). But beyond simple questions of access, language also prevents immigrants from achieving deeper financial literacy, as indicated by a study by the Lutheran Immigration and Refugee Service which showed that even though most non-English speaking immigrants could learn conversational English within 12 to 18 months, it took the average immigrant "as long as 5 years to develop the advanced English skills needed to understand financial contracts" (Lutheran Immigration and Refugee Service & RefugeeWorks, 2005). This same study showed how the language barrier also made the immigrants affected more vulnerable to fraud and financial trouble: multiple cases were found in which non-English speaking immigrants were defrauded by speakers of their native languages whom they trusted to help with personal financial matters (Lutheran Immigration and Refugee Service & RefugeeWorks, 2005).

Studies of immigrant financial behavior in the UK surfaced many of the same insights: lack of English skills are an important barrier to financial access as it increases one's mistrust of financial institutions and the risk of becoming a victim of fraud (Atkinson, 2006; Gibbs, 2010). Gibbs also notes that even migrants from English-speaking countries, the differences in financial terminology can still be confusing (Atkinson & Messy, 2013; Gibbs, 2010).

Both the US and UK-focused studies made similar recommendations to boost financial inclusion among linguistic diverse immigrants – translating of banking materials and publicity, and hiring of multilingual, culturally sensitive staff at bank locations (Gibbs, 2010; United States Government Accountability Office, 2010). Both groups of studies from the US and the UK also cautioned that implementing these recommendations would be costly (Gibbs, 2010; Lutheran Immigration and Refugee Service & RefugeeWorks, 2005; United States Government Accountability Office, 2010). This is reminiscent of the discussion of education in linguistically diverse contexts, in that adapting education to plurilingual populations would be extremely costly, and there is, generally, little motivation for the private sector to meet help produce multilingual educational materials or services without government intervention (Brock-Utne, 2001). There also appears to be little motive for the private sector financial services providers to adapt to linguistic diversity of immigrants in the US or UK. One US government report lists the "private sector" institutions that sponsor financial literacy initiatives for consumers that don't speak English: it includes only a government-owned bank and non-profit financial services industry associations (United States Government Accountability Office, 2010).

2.6.4 Case Studies on the Role of Language in the Business of Financial Inclusion

However, there is some evidence to suggest that private, formal financial services providers may stand to gain by adapting to linguistically diverse contexts. This section will look at three cases in which private financial services providers embraced linguistic diversity as a competitive differentiator and commercial strategy.

One of Guatemala's most successful commercial banks is Banrural, originally founded as a development bank for the rural and micro-enterprise sector in in the country (Oikocredit, 2019). 75% of its loans are made in the country's rural provinces, where most of Guatemala's indigenous population lives (Oikocredit, 2019; Trivelli A & Tarazona S, 2007). Guatemala is linguistically diverse by Latin American standards, due to 40% of its population speaking one of several indigenous Mayan languages (Trivelli A & Tarazona S, 2007). These indigenous customers have traditionally not had access to formal financial services, due to lower rates of education, Spanish literacy, and the fact that many live in isolated rural departments (Furusawa, 2016). Banrural adopted a strategy of placing ATMs in these far-flung, rural departments, co-located with other important institutions such as pharmacies (Chibba, 2009). These ATMs included options for indigenous languages, enabling non-Spanish speakers to use them (Chibba, 2009; Furusawa, 2016). Banrural's embrace of remote, Maya-speaking clients helped it to capture 50% of the market share in rural areas, and made it one of the largest and most well-reputed banks in the country – particularly among its rural client base (Trivelli A & Tarazona S, 2007). Its strategic focus on rural areas and indigenous and female clients has also won the bank praise from the IMF for its effects on financial inclusion in the country, and has helped the bank both to grow and to stabilize its revenues and reduce its risk, thanks to geographic and sectoral diversification (Furusawa, 2016; Trivelli A & Tarazona S, 2007). While it would be an exaggeration to attribute the bank's success *entirely* to embracing Guatemala's linguistic diversity, it is an interesting example of how a small effort (multilingual ATMs) can pay off in numerous ways (client loyalty, diversified risk, higher financial inclusion).

A similar case comes from Equity Bank in Kenya. Equity Bank differentiates itself from other domestic banks in Kenya in that it chooses to "target underserved areas and less privileged households" (Carletti et al., 2012). The authors of one detailed study analysed this strategic approach to determine its effectiveness in increasing financial inclusion in Kenya and, more generally, in Africa. They determine that:

> "…providing financial services to the population segments ignored by traditional commercial banks and generating substantial profits in the process … can be a viable solution to the financial access problem that has hindered the development of the financial sector in many developing countries." (Carletti et al., 2012)

But what does this have to do with language? The authors' methodology uses language in an interesting Instrumental Variable approach. They use the proportion of speakers of minority languages (not English nor Kiswahili) in a district as an instrument for Equity Bank's expansion to that district between 2006 and 2009, to address the endogeneity problem arising from the fact that banks do not expand randomly. The idea behind this instrument is that Equity Bank's target customer demographic in this expansion tended to come from more rural, middle-to-low-income districts. An explicit part of this expansion strategy was to target minority language speakers by always have minority language-speaking associates employed in these new branches. Thus, the proportion of minority language speakers was a good predictor of Equity Bank's decision to open a branch in a district (Carletti et al., 2012). The results of the analyses show that from 2006 to 2009, Equity Bank did indeed adopt a different, more intense branch expansion strategy than peer banks, and that it expanded more than peers in districts with fewer English or Kiswahili speakers. It also shows that the presence of a new Equity Bank branch in a district was significantly correlated with increases in financial inclusion in that district, independent of the actions of other banks, showing that:

> "…the business model of Equity Bank, which targeted middle and low-income segments of the population using such strategies as the local language requirement in its branches, has paid off in terms of greater financial inclusion." (Carletti et al., 2012)

However, this does not necessarily indicate that linguistic diversity and financial inclusion are related in anyway. In fact, the authors of the study specifically warn against making any judgements about language's role in financial inclusion:

> "It is reasonable to argue that the fraction of people from a district speaking a particular language should not be directly linked to whether a particular individual or household has access to a bank account, which is our outcome variable. However, it is important to note that this assumption does not imply that the language an individual speaks is not an important determinant of having a bank account." (Carletti et al., 2012)

Equity Bank's strategy is an example of how formal financial institutions can grow and reduce financial exclusion in linguistically diverse contexts by embracing plurilingualism. The authors are correct however, and in both this case and the case of Banrural it is difficult to determine whether the effects seen were the result of multilingual ATMs and personnel, or of bringing physical access to remote districts, or of other intervening variables.

A more substantial argument for the importance of embracing plurilingualism in financial inclusion efforts come from one last study from Côte d'Ivoire. The qualitative study argues that like performance, companies' language policies – and those of microfinance institutions in particular - are no longer just operational issues, but strategic considerations (Kouame, 2014). Kouame begins his study broadly, analyzing arguments about language and economics, before narrowing down to questions of how language-use choices in Ivoirian microfinance institutions affect interaction with customers. In this broad beginning, he fervently rejects “exoglossia” - the adoption of colonial languages as official languages of nations that were once colonized – and describes how adoption of these languages’ was motivated by a desire for standardization or linguistic normalization, based on views from the Fishman-Pool hypothesis that national development necessitated monolingualism (Kouame, 2014). He argues that an enforced *lingua franca* may hinder domestic development, because insistence on use of the colonial language in business and government will exclude potential producers or consumers that do not speak this language, and will marginalize natural economic actors and activities that were already transacting in local languages (Kouame, 2014).

The more specific second half of the study analyzes the language use of microfinance agents in small cities in Côte d`Ivoire, and interviews agents, customers, and managers about the use of French as opposed to local languages in customer interactions and marketing. It finds that among customers with the lowest levels of education or who are “illiterate” in French, microfinance institutions whose agents are encouraged to engage with customers in local languages are regarded as significantly more trustworthy and empathetic compared to those whose agents use French, which are regarded with feelings of “trouble, antipathy, and difficulty” (Kouame, 2014). Kouame summarizes his study pithily:

> “Though fragmented, this study on the managerial usefulness of linguistic diversity in the Ivoirian microfinance sector demonstrates that the need for plurilingualism as a strategy for winning and retaining customers is now an obvious strategic necessity.” (Kouame, 2014)

None of the cases in this section proves the existence of a relationship between financial inclusion and linguistic diversity. However, all three suggest that embracing plurilingualism may be a winning commercial strategy for banks or microfinance institutions targeting the financially excluded, even in areas of low population density. However, it is useful to recall that the unit of analysis of this study is not the individual bank, client, or community, but an entire country. While these cases seem to indicate the strategic local-scale advantages that plurilingual adaptation can have for firms and for financial inclusion, this tells us little about the country-level effects of linguistic diversity on financial inclusion.

2.7 Literature review summary

Linguistic diversity and financial inclusion are not frequently analysed together in the academic literature on either concept. The few empirical cases that explicitly include both concepts seem to indicate that, at an individual level, language differences – specifically, speaking a language not used in the local, formal financial services industry – leads to voluntary and involuntary financial exclusion (Eves & Titus, 2017; Kouame, 2014; Otieno et al., 2016). This also seems to be true for groups of people as well, such as non-English speaking immigrants in the USA and UK (Atkinson, 2006; Barcellos et al., 2012; Gibbs, 2010; Lutheran Immigration and Refugee Service & RefugeeWorks, 2005), majority-indigenous communities in Latin America (Escalante & Batun, 2018; Furusawa, 2016; Salazar-Cantú et al., 2015), and minority language-speaking communities in Kenya (Carletti et al., 2012).

However, there are no comparative, country-level studies on how these concepts relate. This literature review attempts to paint a picture of such a country-level relationship between the concepts by reviewing how each relates to the more general concept of economic development. It is generally accepted the financial inclusion is positively related to economic development. It is less clear how linguistic diversity relates, but there is generally more evidence that it is negatively related to development. More recent research contests this, indicating that linguistic diversity and economic development are either not related, or even positively related. Interesting case studies also indicate that embracing linguistic plurality can give financial services providers a competitive advantage in linguistic diverse contexts, in addition to boosting financial inclusion (Carletti et al., 2012; Furusawa, 2016; Kouame, 2014).

Several challenges inhibit a potential investigation of the relationship between linguistic diversity and financial inclusion. This review highlights the “broadness” of both concepts, which makes it difficult to define, operationalize, and measure either concept with a single, neat variable. Beyond this, data for measuring linguistic diversity is difficult to find, and data for measuring financial inclusion is almost too broadly available – it is difficult to choose a single variable to use to measure it! Finally, the relationship between these variables may be affected by many intervening and confounding variables. The next section explains how this study attempts to overcome these challenges in its methodology.

Chapter 3: Methodology

3.1 Introduction

This chapter provides the basis for the research design and quantitative approach adopted to investigate the research problem statement. It explains the theoretical basis for the operationalizations of the concepts in this study and lists the data sources used to build them. It discusses the regression approach and estimation technique used to evaluate the potential relationship between linguistic diversity and financial inclusion. It also attempts to justify the approach, while commenting on its weaknesses and the steps taken to mitigate them.

3.2 Research approach

The approach taken in this research is quantitative, explanatory, and cross-sectional in nature. It uses country-level, cross-sectional data, sourced from public or commercially available sources. Though it is explanatory, the research does not aim to establish whether there exists a causal relationship between linguistic diversity and financial inclusion, only whether any significant relationship exists at all. Ideally a panel-based approach with time series data would be preferable to analyse the relationship of these variables over time. However, such data is not easily available for many of the component variables used to build the linguistic diversity variables.

3.3 Research design

To achieve the stated objective of this study, a five-step approach is employed. First, variables are operationalized to represent the concepts of linguistic diversity and financial inclusion. Some of these variables have been previously compiled by other researchers (*BANK*, *ALE*, and *FRN*), and the others will be calculated from data sources discussed in the later in this chapter (*IFI*, *LLG*, *LPC*). Second, the suitability of the variables for regression analysis is assessed, and certain variables are transformed in preparation for analysis. The first and second steps are discussed in section 4.2. Third, a simple correlation analysis is performed to identify any multicollinear variable pairs and to assess high-level relationships between the variables of interest. This is discussed in section 4.3. Fourth, cross-sectional Ordinary Least Squares (OLS) regression models are estimated to examine the relationships between each possible pair of dependent and independent variables, both accounting for and not accounting for any multicollinearity identified in the second step. The relationship between the dependent variable (financial inclusion) and the independent variable (linguistic diversity) is of the greatest interest. However, the models include other control variables – known determinants of country level financial inclusion from the exiting literature – and the results for these controls is briefly evaluated as well. Finally, for each possible combination of a dependent financial inclusion variable and an independent linguistic diversity variable, the best-fit models are combined in a Seemingly Unrelated Regression (SUR) system of equations – one for each dependent financial inclusion variable. These final two steps are discussed in section 4.4.

3.3.1 Data and sample

This study uses secondary data collected from a variety of publicly available data sources, including the World Bank, Ethnologue, the CIA World Factbook, and the United Nations Statistics Division (Agency, 2019; Demirgüç-Kunt et al., 2018; Simons & Fennig, 2018; United Nations Statistics Division, 2018). Data was collected for 243 countries and territories, but only 86 countries had sufficient data to be included in the final analysis. The actual number of observations in the models varies from 61 to 86 countries, due to missing data and the estimation technique. This data is used to construct operationalizations of the two constructs of interest: linguistic diversity and financial inclusion. Country-level linguistic diversity (LD) has been operationalized using four variables, and country-level financial inclusion (FI) has been operationalized using two variables, whose construction and theoretical bases are explained below in section 3.3.3. See Table 3.1 for a summary of the operationalized variables. The data also includes several control variables, which are also country-level statistics from the same sources, primarily from the World Bank's World Development Indicators. See Table 3.1 for a summary of the control variables. See Appendix A for a list of the countries whose data was included in the final analysis.

The period of the data varies slightly. Data used to build the FI variables and the control variables come from 2017 or 2018. In some cases where data for these years were not available in major datasets, such as the World Bank Findex data or UNSD data, additional research was undertaken to find data points from these same years but outside of the original datasets. If this research did not surface such data, values from as early as 2013 were used, where they existed. This approach was only necessary in the case of two control variables: Adult Literacy (*lit*) and the Gini Coefficient (*gini*) (see Table 3.1). This need to use out-of-period data should be considered a weakness of the methodology – future studies may undertake further efforts to estimate more current measures of these variables.

Data used to construct two measures of LD – the largest language group percentage (*LLG)* and the living languages per capita (*LPC)* - came from the years 2009 through 2018, always using the timeliest data available. These variables were calculated with language and population data from Ethnologue and the UNSD. This use of older data is justified because the proportions of language speakers and the number of languages spoken in countries do not necessarily change quickly over time. Indeed, other measures of linguistic diversity base their calculations on broader periods of collected language data, sometimes with data as old as the 1980s (Alesina et al., 2003; Fearon, 2003; Nettle, 2000; Pool, 1972). The *ALE* and *FRN* data are datasets previously compiled by researchers in the early 2000s. This older data could also be considered a weakness of this methodology. Unfortunately, as discussed in the literature review, timely, comprehensive language data is generally unavailable for broad, country-level studies of linguistic diversity (J. L. Arcand, 1996; Nettle, 2000; Posner, 2004).

3.3.2 Empirical model

In step four of the analyses, the ordinary least squares (OLS) regression equation will be used to analyse the relationship between each possible combination of LD and FI variables. FI will be the dependent variable, and LD will be the independent variable, though this is not intended to imply anything about the flow of causality between the variables. Each model will take the standard OLS multiple regression form:

Equation 1: Regression Model

$$FI_i = \beta_0 + \beta_1 LD_i + \beta_2 gdppc_i + \beta_3 lit_i + \beta_4 rpop_i + \beta_5 gini_i + \beta_6 fixed_i + \beta_7 mobile_i + \beta_7 web_i + \beta_8 npl_i + \varepsilon_i$$

Where:

- FI_i is one of the two operationalized variables representing financial inclusion (*BANK*, *IFI*) in a country, indexed over *i*
- *LD* is one of the four operationalized variables representing linguistic diversity (*LLG*, *LPC*, *ALE*, *FRN*) in a country, indexed over *i*
- $gdppc$ is the GDP per capita of a country, measured in current USD

- lit is the adult (15+) literacy rate of a country
- $rpop$ is the percentage of a country's population living in rural areas
- $gini$ is the GINI coefficient of a country, a measure of its income inequality
- $fixed$ is the number of fixed telephone line connections per 100 inhabitants of a country
- $mobile$ is the number of mobile phone subscriptions per 100 inhabitants in a country
- web is the percentage of the population of a country that uses the internet
- npl is the percentage of loans made by a country's banks that are non-performing loans
- ε is an error term that represents the random deviations from the predicted linear relationship

There are eight possible pairs of FI and LD variables, and so eight regressions are run following the form of Equation 1, followed by an additional 16 models accounting for incidences of multicollinearity. All the dependent, independent, and control variables are described in section 3.4 and in the appendices.

In the final step, two SUR models are built using the previous OLS models as inputs. Each SUR model used four of the OLS models – the four most-robust models that have the same dependent financial inclusion variable, but different independent linguistic diversity variables. These models take the form:

Equation 2: SUR Model

$$FI_i = \beta_0 + \beta_i X + \varepsilon_i$$

Where:

- FI_i is a (4x1) vector representing the dependent variables of the four component OLS models. Each row of this vector represents the same financial inclusion variable, either *BANK* or *IFI*
- X is a (nxK_i) matrix whose rows represent n observations common across the four component models, and whose columns represent the K_i columns specific to the i-th component model. The first column of X is a linguistic diversity variable, while the second through K_i-th columns is the control variables used in the i-th component model. "n" will be determined by the number of observations common across all the component

models, and “K” will be 4 for each SUR regression – one component model for each independent linguistic diversity variable

- ε_i is an error term that represents the random deviations from the predicted linear relationship in the i-th component model

3.4 Definition and measurement of variables

This section contains descriptions of the dependent, independent, and control variables used in this study. Summaries of the dependent and independent variables can be found in Table 3.1.

3.4.1 Dependent Variables

The dependent variables are two measures of financial inclusion, which include one “simple” measure, and one more complex index based on the work of Sarma (2008).

3.4.1.1: Percentage of Adults with a Bank Account or Mobile Money Account (*BANK*)

The first dependent variable is the simple percentage of adults with a bank account at a formal financial institution or with a mobile money account. An earlier form of this measurement counted only adults with bank accounts at formal financial institutions, and this measurement has been used by other researchers for its simplicity and easy availability (Wentzel, Diatha, & Yadavalli, 2016). However in recent times, it is criticized for its reductionist view of financial inclusion, particularly at a time when digital technology is transforming financial services (Allen, Demirgüç-Kunt, et al., 2012; Thorsten Beck et al., 2008) As a result, the World Bank Findex data expanded the definition of variable to include ownership of a mobile money account beginning with its 2014 data. (Demirgüç-Kunt et al., 2018).

3.4.1.2: IFI for Traditional Financial Services (*IFI*)

The second dependent variable is an Index of Financial Inclusion (IFI) calculated using a method devised by Sarma (2008). An IFI calculated using this method aggregates different indicators of financial inclusion into a multidimensional index as follows: for each dimension considered, a dimension index (d_i) is calculated:

Equation 3: IFI Dimension Index

$$d_i = \frac{A_i - m_i}{M_i - m_i}$$

In which:

A_i = actual value of dimension i

m_i = minimum value of dimension i

M_i = maximum value of dimension i

The dimension index is then used to calculate the inverse Euclidian distance from the "ideal" IFI – the highest possible measurement of financial inclusion possible for the given dimensions. This gives the IFI for that country:

Equation 4: IFI

$$IFI_i = 1 - \frac{\sqrt{(1 - d_1)^2 + (1 - d_2)^2 + \ldots + (1 - d_n)^2}}{\sqrt{n}}$$

in which n = the number of dimensions included in the IFI. With this tool in mind, *IFI* will measures financial inclusion across countries in a way that reflects more traditional, "brick-and-mortar" financial services, as those provided by a physical bank. This IFI will use four dimensions:

1. To account for banking penetration, the first dimension will be the percentage of the adult population with a bank account at a financial institution.
2. To account for usage of formal financial services, the second dimension will be the percentage of the adult population with a "high frequency" of account use – defined as three or more transactions per month using their formal bank account.
3. To account for access to formal financial services, the third dimension will be number of automated teller machines (ATMs) per 100,000 inhabitants (or per 1,000 square kilometres).
4. To account for credit penetration, the fourth dimension will be the percentage of adults with a loan from a formal financial institution.

All these data points are available through either 2017 Findex data, IMF Financial Access Surveys data, or World Bank Global Payments Systems Survey data (Demirgüç-Kunt et al., 2018; The International Monetary Fund, 2018; The World Bank Group, 2016b).

3.4.2 Independent Variables

The four independent variables are different measures of linguistic diversity and include two "simple" indicators and two ethno-linguistic fractionalization indices (ELFs) compiled by previous researchers.

3.4.2.1: Largest Language Group (*LLG*)

The first independent variable and conceptualization of linguistic diversity is the percentage of a country's population represented by the largest group of native speakers of a language in the country. This data will be compiled by the study author from publicly available data from the United Nations Statistics Division (UNSD) and Ethnologue. To illustrate, according to Botswanan census data from UNSD, the largest language group in Botswana is speakers of Setswana (73% of the total population speak Setswana) (United Nations Statistics Division, 2018). This first measure would be calculated as:

Equation 5: Largest Language Group %

$$LLG = 1 - LLG\%$$

Thus, for Botswana, *LLG* would be 1 – 0.73, or 0.27. The *LLG%* is subtracted from one because this orients the measure so that larger values indicate higher linguistic diversity. The other variables for linguistic diversity are naturally oriented this way, so this is done to preserve consistency. The advantage of this measure is its simplicity. It only depends on a measure of the total population of a country, and the population that speak the various languages in that country (Fishman, 1966; Nettle, 2000; Pool, 1972).

However, *LLG* doesn't adequately capture some aspects of linguistic diversity, such as multilingualism, and so it tends to overestimate linguistic diversity, as discussed in the literature review (J. L. Arcand, 1996; Nettle, 2000). To mitigate this drawback, the variable is not calculated as the population percentage of largest group of native speakers, but the population percentage of the largest group of speakers – native or not. This better approximates the proportion of the population able to communicate in a single language, regardless of their mother tongue. A further disadvantage of this measure is that there is no consistent period for the language census data on which it is based. Country-level data on the numbers of speakers of languages in countries varies in its timeliness. Of the 242 countries collected, only 24 have data from 2017 or 2018. However, as explained in section 3.3.1, this study will make use of the countries whose language data come from the ten most recently available years, 2009 through 2018, totalling 172 country observations. This approach to using broad cross-sections of language data across data has been used in the construction of other measures of linguistic diversity, including by Alesina et al. and Fearon – whose indices are *ALE* and *FRN* in this study (Alesina et al., 2003; Fearon, 2003).

3.4.2.2: Languages Per Capita (*LPC*)

The second independent variable will be the "languages per capita" – that is, the number of languages spoken natively in a country divided by the population in millions, as such:

Equation 6: Languages per Capita

$$LPC = \frac{\#\ Living\ Languages}{\left(\frac{Population}{1{,}000{,}000}\right)}$$

This data will also be compiled by the study author from UNSD and Ethnologue. The number of languages will be determined as those between levels 1 and 6a on the Expanded Graded Intergenerational Disruption Scale (EGIDS), a scale developed by Ethnologue to measure language development (Simons & Fennig, 2018). Filtering by EGIDS level will avoid counting languages that are either no longer in active use by people today, and therefore not relevant to the research question and hypothesis. Ethnologue has the required data for 232 countries and dependencies, and UNSD has recent population figures for all of these.

The advantage of this measure, devised by Daniel Nettle, is that it helps reduce overestimation of diversity for larger or more populous countries, and underestimation of the same figures for smaller or less populous countries (Nettle, 2000).

Because this data is based on the count of living languages, and not on the populations of speakers, this measure does not face the same problems with timeliness that LLG faces. The population figure used is for the country's full population, regardless of spoken language, which is substantially easier to obtain as well.

One key drawback is that this measurement leads to massively inflated figures for small, sparsely populated countries. For this reason, any countries with less than 10,000 km^2 in total area or less than 100,000 in total population have been dropped from analysis, leaving a total 212 country-level observations in the data. This engenders a further disadvantage, namely that this measure excludes many small, island nations, such as the Cook Islands, the Maldives, and the Federated States of Micronesia, and so the study's results may not be applicable or relevant to those countries' contexts.

3.4.2.3: Alesina's Ethnolinguistic Fractionalization Index (*ALE*)

The third independent variable is an ethnolinguistic fractionalization index (ELF) previously compiled by Alesina et al., and discussed in more detail earlier in section 2.5 (Alesina et al., 2003). The measure is calculated:

Equation 7: Alesina ELF

$$ALE_j = 1 - \sum_{i=1}^{N} {p_{ij}}^2$$

where p_{ij} is the population share of language group i (i = 1, …, N) in country j. This is a basic Greenberg or Herfindahl index, with the data painstakingly collected by Alesina and his co-researchers (Greenberg, 1956; Herfindahl, 1950). This data covers 201 territories or dependencies.

3.4.2.4: Fearon Linguistic Diversity Index (*FRN*)

The last independent variable is Fearon's cultural diversity index, which is also an ELF produced through a similar approach to Alesina et al, but which also accounts for the similarity of languages spoken within a country (Fearon, 2003). The reason for this builds on Greenberg's "expectation": if different languages are similar, this similarity should mitigate the miscommunication problems they engender, as they may be mutually intelligible to some degree (Greenberg, 1956). To approximate this effect, Fearon subtracts a "resemblance factor" from overall diversity index figures. This factor takes values between 0 and 1, with 0 signifying maximum difference between a countries' languages, and 1 signifying maximum similarity. Thus, countries with more similar languages have their diversity index "discounted" by a resemblance factor closer to 1, lowering their overall diversity score (Fearon, 2003; Greenberg, 1956). These resemblance factors are based on phylogenetic language tree diagrams, and are also available from Ethnologue (Fearon, 2003; Simons & Fennig, 2018). The data itself is shared in Fearon's paper – this study makes use of the data series labelled "C", which he calls "cultural fractionalization" (Fearon, 2003).

3.4.3 Control Variables

Eight control variables are used in the analysis, adopted from Sarma's cross-country analysis of the determinants of country-level financial inclusion. These controls include socio-economic variables, infrastructure related variables, and one banking sector related variable, and are found to be significant determinants of financial inclusion at a country level – both in this and other empirical studies (Sarma & Pais, 2010b). Their inclusion helps assure that any observed relationship between the financial inclusion and linguistic diversity variables is not spurious.

3.4.3.1: GDP per capita (current USD) (*gdppc*)

Of the socio-economic explanatory variables considered by Sarma et al (2010), country-level GDPPC is the most significant ($p < 1\%$), and is positively related to levels of financial inclusion in their study. This significant, positive relationship exists at the level of individuals and households as well. Studies based on survey data demonstrate that higher individual income levels (which GDP purports to measure in aggregate) are important determinants of financial inclusion in Sub-Saharan Africa (Chikalipah, 2017; Zins & Weill, 2016b), and more specifically in Ghana and Kenya (Akudugu, 2013; Johnson & Arnold, 2012). The same is true for income measured at the level of households and municipalities in Mexico (Salazar-Cantú et al., 2015). All these studies found that financial inclusion is positively related to measures of country-level, household-level, or individual-level income. Interestingly, the studies all arrive at this result despite different areas of focus. For example, Johnson et al focus on consumer access to financial markets, Zins and Weill focus on formal, informal, and digital consumer financial services (Johnson & Arnold, 2012; Zins & Weill, 2016b). Given the unanimity of results across these studies, it is expected that the relationship of the dependent variables to this control variable will be positive.

3.4.3.2: Adult (15+) literacy rate (*lit*)

The relationship between literacy and financial inclusion has been thoroughly studied in a variety of contexts. Several country-level studies have found that higher levels of individual literacy correspond to higher levels of financial inclusion, including in Ghana, Kenya, Mexico, and Côte d'Ivoire (Akudugu, 2013; Johnson & Arnold, 2012; Kouame, 2014). Sarma identifies higher adult literacy rates as one of the key determinants of country-level financial inclusion (Sarma & Pais, 2010). An OECD report explicitly links linguistic diversity to financial inclusion: it determined that higher literacy rates enable potential customers to learn about and use financial services, and suggests that this would be more difficult in linguistically diverse countries or contexts (Atkinson & Messy, 2013).

As discussed in the literature review, even when literacy is not an explicitly considered factor in empirical studies of barriers to financial inclusion, it often figures in researchers' recommendations. Take, for example, Chaibou and Bihari's recommendations to localize language of promotional materials and ATMs in Niger and India, respectively, to assist excluded consumers that are "illiterate" in French or English (Bihari, 2011; Chaibou, 2019). For these reasons, it was expected that this control variable would also have a positive relationship with this study's dependent variables.

3.4.3.3: Rural population as a % of the total population (*rpop*)

The literature on the relationship between a population's rurality and its financial inclusion is more mixed in its results. Leyshon and Thrift documented a "flight to quality" by formal financial institutions in the UK and USA, in which banks closed high-cost rural branches to pursue growing, increasingly urban middle and upper class customers – effectively excluding rural customers from the financial system (Leyshon & Thrift, 1995). Their study pointed out that providing formal financial services to less dense, less wealthy rural populations is not always commercially viable. Indeed, broader data and studies from the World Bank and OECD do indeed suggest that rural individuals and households have higher probabilities of being financially excluded across geographies (Atkinson & Messy, 2013; Demirgüç-Kunt et al., 2018). Single country studies have also suggested that physical proximity to providers of financial services is important to increasing financial inclusion, and that a larger rural population makes it difficult to achieve this proximity for that population. This appears to hold true for consumers from the more rural north of Uganda compared to those in its more urban centre (Katoroogo, 2016). The same is true of formal (bank-mediated) financial access in Kenya: rural Kenyan consumers are far less likely to be financially included via traditional bank than their urban counterparts (Johnson & Arnold, 2012).

However other studies have suggested that rurality and financial inclusion are not related. Chikalipah showed that population density, a proxy for rurality, was not related to financial inclusion in Sub-Saharan African countries – offering low-density, high-inclusion Botswana and high-density, low-inclusion Malawi as examples of where this relationship seemingly reverses (Chikalipah, 2017). Wentzel et al's study arrived at similar results for bottom-of-the-pyramid consumers in South Africa, suggesting that true determinants of individual financial exclusion, such as unemployment and lack of education, are simply more prevalent in rural areas (Wentzel, Diatha, Seshachal, et al., 2016).

Despite the mixed empirical stances on the relationship between financial inclusion and rurality of a population, it was expected that the relationship between this control variable and the study's dependent variables would be negative.

3.4.3.4: GINI coefficient (*gini*)

The GINI coefficient is a measure population's deviation from perfect income inequality, with 0 representing perfect equality, and 1 representing perfect inequality. Thus, this control variable captures the degree of income inequality in a country. Countries' level of income inequality have been shown to be negatively related to financial inclusion – that is, higher inequality corresponds to lower financial inclusion (Kempson et al., 2004; Sarma & Pais, 2010). Interestingly, some researchers have shown that this relationship between financial inclusion and income inequality is causal in nature, and that increasing the former can decrease that latter – both in general, country-level contexts (García-Herrero & Martínez Turégano, 2015) as well as in more granular contexts, such as Salazar-Cantú et al's study of the phenomenon in Mexican municipalities (Salazar-Cantú et al., 2015). Given the apparent consensus on the negative nature of the relationship between income inequality – as measured by the GINI coefficient – and financial inclusion, it is expected that this study's dependent variables will have a negative relationship with this control variable.

3.4.3.5: Fixed telephone connections (*fixed*) *and* Mobile telephone subscriptions (*mobile*) *and* % of the population that uses the internet (*web*)

These three variables are the number of fixed telephone connections per 100,000 residents of a country, the number of mobile phone subscriptions per 100,000 residents of a country, and the percentage of a country's population that uses the internet. These three control variables fall under the umbrella of "infrastructure" variables in Sarma et al's study on the country-level determinants of financial inclusion (Sarma & Pais, 2010). The study showed that all three have positive, significant associations with higher levels of financial inclusion, which the authors suggest may “indicate that connectivity and information play an important role in financial inclusion” (Sarma & Pais, 2010). World Bank researchers confirmed this as well, determining that physical infrastructure in general, and telephone connections per capita in particular, makes formal financial service delivery more efficient and, by extension, more generally accessible (T. Beck et al., 2007). Allen et al suggest this is already happening in Kenya, thanks to the success of M-PESA – though the financial products it offered couldn’t satisfy all consumers’ or firms’ needs, at least at the time of writing (Allen, Carletti, et al., 2012).

This capability – both realized and potential - of increased mobile phone penetration and increased access to the internet to increase financial inclusion is a subject of frequent study, especially in Sub-Saharan Africa (SSonko, 2010). Many developing countries with lower levels of formal financial inclusion demonstrate much higher adoption of digital financial services than their developed counterparts (Naghavi, 2019). Increased penetration of mobile phones and internet access increases access to new financial services, as well as digital forms of traditional financial services – both formal, such as bank accounts and loans, and informal, such as traditional savings groups (Naghavi, 2019).

However, Kempson et al (2004) note a mixed effect of increased mobile phone adoption and internet use in developed economies – as these technologies penetrate countries, they can accelerate closure of high-cost physical bank branches in rural or low-income urban areas, which reduces access to more traditional financial services for these communities. Despite this, it is expected that the two dependent variables which measure more traditional financial inclusion (*BANK* and *IFI*) will have positive relationships with these control variables.

3.4.3.7: Non-performing loans (*npl*)

The final control variable relates to the health and soundness of countries’ banking sectors. It is the percentage of loans granted by a country’s banks that are non-performing loans (NPLs) – that is, loans which have or are expected to default. Though many would expect a higher proportion of NPLs if banks lend to poorer, financially excluded people who subsequently default, Sarma et al’s study demonstrated that in fact the opposite is true (2010). Studies specifically focused on the relationship between financial inclusion and financial stability have shown that greater financial inclusion leads to greater stability – particularly when stability is measured using non-performing loans or assets (Morgan & Pontines, 2014).

Chen et al confirm this result at national and sub-national levels in the Chinese context, showing that greater financial inclusion had an inhibitory effect on the proportion of NPLs in regional banks in the PRC – particularly in regions in which the base level of financial inclusion was already comparatively low (Chen et al., 2018). The apparent salutary effect of including previously-excluded customers on banks' balance sheets is reminiscent of the Equity Bank and Banrural cases in Kenya and Guatemala, discussed in the literature review (Carletti et al., 2012; Furusawa, 2016; Trivelli A & Tarazona S, 2007). However, there unfortunately are not data on the Equity Bank's or Banrural's NPLs from either study. The idea of studying bank-level financial stability as it relates to financial inclusion would be an interesting path for further study.

Based on the empirical work briefly detailed above, it is expected that this control variable will have a negative relationship with the dependent variables – that is, higher proportions of non-performing loans in a country will correlate to lower measures of financial inclusion.

Table 3.1: Measurement and sources of variables

Operationalized Variables		
Name	**Description**	**Source**
LLG – Largest Language Group	The percentage of the country's population comprised by the largest group of fluent speakers of a single language. – From Ethnologue, the CIA World Factbook, and UNSD	Ethnologue (otherwise UN Statistics Division)
LPC – Languages per Capita	The number of "living" languages per 1 million inhabitants of a country: the number of languages between 1 and 6a on the EGIDS Scale, divided by population in millions. Excludes countries smaller than 10,000 square kilometers or with fewer than 100,000 people in population. – from Ethnologue, the CIA World Factbook, and UNSD	Ethnologue (otherwise UN Statistics Division)
ALE – Alesina et. Al ELF Index	An "ethnolinguistic fractionalization" index built by Alesina et al., a simple Herfindahl index.	Alesina et al, 2003
FRN – Fearon Linguistic Diversity Index	Another Herfindahl index for linguistic diversity that attempts to account for structural diversity between languages.	Fearon, 2003
BANK - % of Adults with a Bank Account	The percentage of a country's adult population (those age 15 and above) that own a bank account at a formal financial institution or with a mobile money provider. – From the World Bank Findex	World Bank Findex
IFI – Traditional Banking Sarma IFI	An Index of Financial Inclusion of the type first suggested by Sarma (2008) using four components that approximate demand-side financial inclusion indicators for financial services. – From the World Bank Findex and IMF Financial Access Survey	World Bank Findex and IMF Financial Access Survey
Control Variables		
gdppc - GDP per capita (current USD)	GDP per capita is gross domestic product divided by midyear population.	World Bank, OECD
lit - Literacy rate	Adult literacy rate is the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life.	UNESCO Institute of Statistics
rpop – Rural population	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population, taken as a percentage of the total population.	World Bank
gini – GINI Index	Gini index measures the extent to which the distribution of income among individuals or households within an economy deviates from a perfectly equal distribution. Thus, a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.	World Bank
fixed -Fixed telephone subscriptions per 100 people	"Fixed telephone subscriptions" refers to the sum of active number of analogue fixed telephone lines, voice-over-IP (VoIP) subscriptions, fixed wireless local loop (WLL) subscriptions, ISDN voice-channel equivalents and fixed public payphones.	World Bank, ITU
mobile – Mobile telephone subscriptions per 100 people	Mobile cellular telephone subscriptions are subscriptions to a public mobile telephone service that provide access to the PSTN using cellular technology.	World Bank, ITU
web - % of population using the internet	Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc. – from the ITU and World Bank	World Bank, ITU
npl – % non-performing loans to total gross bank loans	Bank nonperforming loans to total gross loans are the value of nonperforming loans divided by the total value of the loan portfolio (including nonperforming loans before the deduction of specific loan-loss provisions). The loan amount recorded as nonperforming should be the gross value of the loan as recorded on the balance sheet, not just the amount that is overdue.	IMF

3.5 Estimation Technique

This study estimates eight different OLS regression models, one for each possible pair of two dependent, financial inclusion variables with four independent, linguistic diversity variables. The OLS technique relies on several key assumptions in order to produce reliable, unbiased estimates. These include Normality, homoskedasticity, and lack of correlation in the error terms, and lack of correlation in the independent variables. (Casson & Farmer, 2014; Nau, 2014) A series of tests are employed to ensure the models created meet these assumptions: the F-test is used to evaluate the overall model significance, the Breusch-Pagan Lagrange Multiplier to evaluate the homoskedasticity of the error terms (Breusch & Pagan, 1979), and the Jarque-Bera test to evaluate the Normality of the error terms (Jarque & Bera, 1980). The correlation analysis in step three of the analysis is used to identify multicollinear independent variables, and a stepwise re-estimation is done to account for these. This produces several models for each possible pair of financial inclusion and linguistic diversity variables, from which the most robust model for each of the eight pairs is chosen for use in the next step.

Subsequently the four OLS models chosen for each dependent variable are re-estimated using a Seemingly Unrelated Regression (SUR) model for each of the two dependent financial inclusion variables. The SUR technique is appropriate in this context, as it is useful for analyses involving multiple linear models of phenomena that are believed to be interdependent or related. This relation manifests itself in correlation between the error terms across the component models (Ulwodi, 2017; Zellner & Lee, 1962). SUR models are useful econometric tools for joint estimation of multiple equations that describe a single model, which appear to be "seemingly unrelated", but which characterize common aspects of a single phenomenon, and thus are likely to have correlated error terms across component equations (Fiebig, 2003; Goedecke et al., 2018; Ulwodi, 2017; Zellner & Lee, 1962) The models built in the last section are good candidates for SUR modelling, since all the models attempt to describe the same relationship – linguistic diversity and financial inclusion at a country level – using different variables. Note that the same variables may use in component models – but if all the independent variables are exactly the same, SUR "collapses" to simple OLS regression for each dependent variable (Fiebig, 2003; Zellner & Lee, 1962). In this study's approach, the same dependent variable, representing financial inclusion, is conserved across the component OLS models of each SUR analysis, while the linguistic diversity independent variable and the control variables employed change across the input equations.

Chapter 4: Discussion of Findings

4.1 Introduction

This chapter details the steps and findings of the analyses described in the previous chapter. It begins with descriptive statistics of the data, followed by a correlation analysis, and ends with the construction of various regression models as outlined in the previous chapter.

4.2 Data Exploration and Preparation

After collecting the data and operationalizing the financial inclusion and linguistic diversity variables in the manner described in sections 3.4.1 and 3.4.2, it was noted that some variables contained extreme outliers, and not all the variables had similar ranges of values. Since a regression model's coefficients are sensitive to the scales of its inputs (Keppel, 1991), it was determined that these factors could reduce the interpretability of subsequent regression results. To identify and eliminate outliers, all variables were standardized to standard Normal "z-score" values, and any observations outside of three standard deviations from zero were dropped. Then all variables not already in a range of zero to one where standardized to this range, yielding a dataset of values between 0 and 1 for all variables.

***Table 4.1** – Summary statistics*

	Mean	Std Dev.	Min.	25%	50%	75%	Max.	N
BANK	0.59	0.27	0.06	0.37	0.56	0.83	0.99	157
IFI	0.29	0.17	0.005	0.16	0.26	0.4	0.78	153
LLG	0.24	0.22	0.00	0.05	0.16	0.39	0.87	235
LPC	0.07	0.13	0.00	0.014	0.03	0.09	1.00	178
ALE	0.36	0.29	0.00	0.09	0.32	0.61	0.92	212
FRN	0.31	0.21	0.00	0.14	0.29	0.49	0.73	154
GDPPC	0.17	0.23	0.00	0.02	0.07	0.22	1.00	198
LIT	0.84	0.18	0.32	0.75	0.93	0.98	1.00	140
RPOP	0.39	0.24	0.00	0.2	0.38	0.58	0.87	213
GINI	0.39	0.08	0.25	0.33	0.37	0.43	0.59	122
FIXED	0.16	0.16	0.00	0.02	0.13	0.24	0.59	200
MOBILE	0.47	0.18	0.00	0.38	0.48	0.59	1.00	202
WEB	0.56	0.29	0.00	0.3	0.61	0.8	1.00	206
NPL	0.06	0.06	0.00	0.02	0.04	0.08	0.29	128

Note: BANK = % of population with a bank or mobile money account; IFI = IFI for Traditional Financial Services; LLG = Largest Language Group; LLC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index, compiled with Python, author's elaboration (The pandas development team, 2020)

It is worth noting that higher values for the financial inclusion and linguistic diversity variables denote higher amounts of financial inclusion and linguistic diversity, respectively. The summary statistics of the variables after this cleaning are presented above in Table 4.1.

Nearly all the measures of the linguistic diversity variables are right-skewed, with some large outliers and the means closer to the minimum side of the range of values. This indicates most countries tend to have lower measures of linguistic diversity, with some outliers that are highly diverse. The measures of financial inclusion, on the other hand, have less skewed distributions. *IFI* approximates a (non-standard) Normal distribution, though is still slightly right skewed.

4.3 Correlation Analysis

As the third step of the analysis, Pearson correlation coefficients were calculated for each pair of variables. These coefficients serve three purposes in our analysis: First, they serve as an informal test of criterion validity. Variables that measure the same concept should have strong, positive correlations. If the linguistic diversity variables exhibit strong positive correlations to one another, this would indicate that the operationalized variables are measuring similar outcomes despite their different construction methods and data sources. The same is true of the financial inclusion variables. Second, they give an initial indication of the strength and direction of the relationship between each LD-FI variable pair, which can be tested further in the subsequent regression analyses (Gravetter & Wallnau, 2013; Hair et al., 2018). For instance, a significant, negative coefficient may suggest that the two variables, and therefore the two concepts, are negatively related. Finally, correlation analysis helps to identify collinear independent variables before a regression analysis. To address the first two purposes, a table of the coefficients for the LD-FI pairs is displayed in Table 4.3 below.

Positive, significant coefficients between variable pairs meant to operationalize the same concept, such as *LLG-ALE* and *BANK-IFI*, suggest that both the different linguistic diversity variables and the different financial inclusion variables exhibit criterion validity. At the same time, the table shows that all variable pairs between one financial inclusion variable and one linguistic diversity variable have weak negative correlations, suggesting that they have negative relationships. This suggests that the relationship between financial inclusion and linguistic diversity is comparable to the "Fishman-Pool" interpretation of the relationship between economic development and linguistic diversity.

Table 4.2 - *Correlation Coefficients for Variable Pairs of Interest*

	BANK	IFI	LLG	LPC	ALE	FRN
BANK	1					
IFI	0.91	1				
LLG	(0.33)	(0.39)	1			
LPC	(0.17)	(0.19)	0.13	1		
ALE	(0.25)	(0.32)	0.61	0.17	1	
FRN	(0.19)	(0.24)	0.51	0.37	0.73	1

Note: BANK = % of population with a bank or mobile money account; IFI = IFI for Traditional Financial Services; LLG = Largest Language Group; LLC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; Calculated with Python, author's elaboration, (The pandas development team, 2020)

The third and most important purpose of correlation analysis in this study is to test for multicollinearity between the independent variables. Absence of multicollinearity is an important assumption of OLS regression analysis, and large, non-zero coefficients of correlation between independent variables would indicate that this assumption is violated. The correlation coefficients between all the independent variables to be used in the regression analysis are shown below in Table 4.3.

The bolded coefficients indicate independent variable pairs with Pearson correlation coefficients with absolute value greater than 0.7. According to Kennedy (2008), regressors that are correlated to this extent could bias regression results due to multicollinearity. Among these variables, there are three such cases: *ALE* and *FRN* (0.73), *lit* and *web* (0.7), and *fixed* and *web* (0.73). The first pair presents no problem, as *ALE* and *FRN* will only ever be used in separate OLS models. The multicollinearity of *lit* and *fixed* with *web* will be addressed in a step-wise approach in the next section, in which each model is re-estimated using *lit* and *fixed*, but not *web* as control variables, and then re-estimated again using *web* but not *lit* nor *fixed* as control variables. One potential weakness of this study is that many of the other control variables pairs also have strong correlations - some with just under 0.7 as their coefficient, such as *gdppc* and *web* with 0.69, or *rpop* and *web* with (0.66).

The signs of the correlation coefficients between any of the linguistic diversity variables and any of the control variables seem to follow the intuition of the "Fishman-Pool" hypothesis. All four linguistic diversity variables are positively correlated with *rpop*, *gini*, and *npl*, and negatively correlated with *gdppc*, *lit*, *fixed*, *mobile*, and *web*. None of these correlations are strongly negative (≤ (0.5)) but this consistency suggests that linguistic diversity does indeed have a negative relationship with different measures of national socioeconomic development, as postulated by Greenberg and later studied by Fishman, Pool, and others (Fishman, 1966; Greenberg, 1956; Nettle, 2000; Pool, 1972).

Table 4.3 *- Correlation Coefficients for Independent Variable Pairs*

	LLG	LPC	ALE	FRN	gdppc	lit	rpop	gini	fixed	mobile	web	npl
LLG	1											
LPC	0.13	1										
ALE	0.61	0.17	1									
FRN	0.51	0.37	**0.73**	1								
GDPPC	(0.25)	(0.13)	(0.26)	(0.22)	1							
LIT	(0.45)	(0.08)	(0.49)	(0.35)	0.40	1						
RPOP	0.31	0.10	0.25	0.29	0.29	(0.45)	1					
GINI	0.12	0.15	0.10	0.03	(0.41)	(0.04)	0.16	1				
FIXED	(0.44)	(0.15)	(0.39)	(0.31)	0.67	0.58	(0.54)	(0.38)	1			
MOBILE	(0.20)	(0.06)	(0.16)	(0.07)	0.38	0.48	(0.36)	(0.20)	0.42	1		
WEB	(0.41)	(0.11)	(0.36)	(0.31)	0.69	**0.70**	(0.66)	(0.37)	**0.73**	0.60	1	
NPL	0.30	0.23	0.30	0.25	(0.36)	(0.47)	0.35	0.15	(0.34)	(0.36)	(0.45)	1

Note: LLG = Largest Language Group; LPC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; gdppc= GDP per capita; lit = Adult literacy %; rpop = Rural population %; gini = Gini Coefficient; fixed = Fixed telephones per 100 inhabitants; mobile = Mobile telephones per 100 inhabitants; web = % population using internet; npl = Non-Performing Loan %; Calculated with Python and the pandas library, author's elaboration; (The pandas development team, 2020). ***Bold text*** *indicates a correlation coefficient with absolute value > 0.7, which may suggest that the two variables are collinear, violating the multicollinearity assumption of linear regression.*

4.4 Regression Results

This section discusses the results of the regression analyses, with focus on the coefficients of determination of the linguistic diversity independent variables, their significance, and what this implies about the relationship between financial inclusion and linguistic diversity. It also looks at the coefficients of the control variables and their significance. As discussed in the previous chapter, this analysis begins with estimation of Ordinary Least Squares (OLS) models for the financial inclusion variables using different combinations of the linguistic diversity variables and other controls, followed by re-estimation of the models using Seemingly Unrelated Regression (SUR).

4.4.1 Initial OLS Regression Models

The initial results of the regression coefficients without accounting for multicollinearity are presented in Table 4.5, with four initial models for the dependent variable *BANK* on the left, and four initial models for the dependent variable *IFI* on the right. While remaining conscientious that these initial models exhibit multicollinearity, the analysis begins by checking these models' conformity with the assumptions of OLS regression. The results show that none of the eight models returned significant ($p < 10\%$) Jarque-Bera (JB) test statistics. The null hypothesis of the JB test is that the skewness and kurtosis of the error terms of the regression are both zero – as they would be if they came from a Normal distribution (Hall & Lilien, 2014; Jarque & Bera, 1980). This indicates that we cannot reject this null hypothesis, and that the assumption regarding Normality of the error terms is not violated. Similarly, the Breusch-Pagan (BP) test statistics, whose null hypothesis is that the error terms are homoscedastic, are also not significant at 10% for all the models, indicating that the assumption regarding heteroskedasticity of the error terms is also not violated (Breusch & Pagan, 1979; Hall & Lilien, 2014). The F-test statistic is highly significant for all the models. The null hypothesis of this test is that all of the coefficients of the model are zero – essentially testing for any significance of the regression coefficients whatsoever (Hall & Lilien, 2014). The highly significant p-values of this test for all eight models indicates that this null can be rejected, and that variance of at least one of the chosen regressors is significantly explanatory of variance of the dependent financial inclusion variables. The R^2 for the four models of *BANK* varies between 0.55 and 0.61, while for the four models of *IFI* varies between 0.68 and 0.71.

The coefficient of three proxies (*LLG*, *ALE*, *FRN*) of linguistic diversity variables are positive at significant at 5% in in the *BANK* model while only *FRN* is observed to be positive and significant at 10% in *IFI* model. This suggests that linguistic diversity enhances financial inclusion in the sample countries. Based on the Fishman-Pool interpretation of the relation between linguistic diversity and economic development, and the positive relationship between economic development and financial inclusion, this apparent conclusion seems counterintuitive. A positive, or at least ambiguous relation between linguistic diversity and economic development (J. Arcand & Grin, 2013) or between linguistic diversity and financial inclusion (Allen, Carletti, et al., 2012; Kouame, 2014) has been identified in more granular studies, but always under specific empirical circumstances, and never at a cross-sectional, country level. Indeed, most studies of linguistic diversity and economic diversity at a country level determine that the two variables are negatively related (Fishman, 1964; Nettle, 2000; Pool, 1972; Posner, 2004), and no country level, cross sectional studies of linguistic diversity and financial inclusion exist.

Table 4.4 – *Eight initial regression models for dependent variable* BANK, IFI

	Models with *BANK* as Dependent Variable								**Models with *IFI* as Dependent Variable**							
	Coeff.	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*
Intercept	(0.26) 0.32	(0.82)	(0.21) 0.34	(0.62)	(0.26) 0.32	(0.83)	(0.24) 0.33	(0.74)	(0.31)* * 0.15	(2.11)	(0.25) 0.15	(1.67)	(0.30)** 0.15	(2.02)	(0.26)* 0.15	(1.76)
LLG	**0.21** 0.10**	**2.15**	-	-	-	-	-	-	**0.08 0.05**	**1.67**	-	-	-	-	-	-
LPC	-	-	**0.30 0.33**	**0.91**	-	-	-	-	-	-	**(0.08) 0.15**	**(0.53)**	-	-	-	-
ALE	-	-	-	-	**0.19** 0.08**	**2.47**	-	-	-	-	-	-	**0.04 0.04**	**1.17**	-	-
FRN	-	-	-	-	-	-	**0.22** 0.10**	**2.27**	-	-	-	-	-	-	**0.08* 0.04**	**1.77**
GDPPC	1.00*** 0.32	3.15	1.04*** 0.33	3.13	1.00*** 0.31	3.20	0.99*** 0.32	3.12	0.46*** 0.14	3.23	0.47*** 0.15	3.18	0.47*** 0.15	3.22	0.46*** 0.14	3.26
LIT	0.49 0.33	1.48	0.42 0.35	1.22	0.41 0.32	1.27	0.41 0.33	1.25	0.40*** 0.15	2.72	0.13** 0.15	2.07	0.37** 0.15	3.22	0.36** 0.15	2.46
RPOP	0.18 0.14	1.33	0.23 0.14	1.61	0.18 0.13	1.34	0.18 0.14	1.27	0.09 0.06	1.52	0.11* 0.06	1.70	0.10 0.06	1.61	0.08 0.06	1.23
GINI	0.22 0.30	0.74	0.15 0.32	0.47	0.23 0.30	0.79	0.25 0.31	0.80	0.15 0.13	1.10	0.15 0.14	1.07	0.15 0.14	1.09	0.14 0.14	0.97
FIXED	0.59** 0.26	2.24	0.45 0.29	1.56	0.58** 0.26	2.26	0.52* 0.28	1.83	0.24** 0.12	2.00	0.22* 0.13	1.77	0.22* 0.12	1.81	0.25* 0.13	1.99
MOBILE	0.10 0.16	0.63	0.14 0.16	0.86	0.14 0.15	0.90	0.10 (0.16)	0.62	0.09 0.07	1.21	0.12 0.07	1.59	0.10 0.07	1.45	0.08 0.07	1.15
WEB	(0.05) 0.18	(0.26)	0.00 0.20	0.00	0.00 0.18	(0.02)	0.01 0.18	0.03	(0.01) 0.08	(0.11)	0.02 0.09	0.25	0.01 0.08	0.16	0.00 0.08	(0.05)
NPL	(0.27) 0.53	(0.51)	(0.01) 0.54	(0.02)	(0.28) 0.52	(0.54)	(0.04) 0.51	(0.09)	(0.26) 0.24	(1.08)	(0.13) 0.24	(0.55)	(0.21) 0.24	(0.89)	(0.21) 0.23	(0.94)
R^2	0.60		0.55		0.61		0.58		0.70		0.68		0.69		0.71	
N	63		62		63		61		63		62		63		61	
	Test stat	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*
JB Test	1.96	0.38	0.57	0.75	1.56	0.46	1.65	0.44	1.92	0.38	0.86	0.65	1.10	0.58	0.32	0.85
BP Test	13.7	0.13	9.63	0.38	10.9	0.28	9.29	0.41	13.7	0.13	8.89	0.45	9.47	0.40	7.73	0.56
F Test	8.80***	0.00	6.95***	0.00	9.17***	0.00	7.82***	0.00	13.6***	0.00	12.5***	0.00	13.2***	0.00	13.6***	0.00

Note: LLG = Largest Language Group; LPC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; gdppc= GDP per capita; lit = Adult literacy %; rpop = Rural population %; gini = Gini Coefficient; fixed = Fixed telephones per 100 inhabitants; mobile = Mobile telephones per 100 inhabitants; web = % population using internet; npl = Non-Performing Loan %; JB = Jarque-Bera Test; BP = Breusch-Pagan Test; Calculated with Python, author's elaboration; (Seabold & Perktold, 2010). **** denotes significance at p = 1%, ** at p = 5%, * at p = 10%;*

However, the positive relationship between financial inclusion and linguistic diversity suggested by these models' coefficients is not consistently significant across all combinations of financial inclusion and linguistic diversity variables.

With regard to the expected relationships between the dependent variable and control variables described in the previous chapter, only about half of these expectations were realized in these eight initial models. National income, as measured by GDPPC in current US dollars (*gdppc*) had a highly significant ($p < 1\%$), positive relationship with both *BANK* and *IFI* across all eight models. This supports the results of previous studies of national income as a determinant of country level financial inclusion (Chikalipah, 2017; Sarma & Pais, 2010b; Zins & Weill, 2016a). Adult literacy (*lit*) also had a significant ($p < 5\%$) positive relationship with *IFI*, and an insignificant positive relationship with *BANK*. This is in line with similar results from the same studies, as well as many single-country studies of the determinants of financial inclusion (Akudugu, 2013; Bihari, 2011; Chaibou, 2019a; Johnson & Arnold, 2012; Kouame, 2014). Finally, the rate of non-performance of a country's banks' loans (*npl*) had a negative but insignificant relationship across the models, which aligns with previous studies of banking sector health and its effects on financial inclusion (Chen et al., 2018; Morgan & Pontines, 2014).

Fixed telephone connections per capita (*fixed*) was also positively related to *BANK* and *IFI*, and often but not always significant. Curiously, mobile telephone subscriptions per capita (*mobile*) and the percentage of the population actively using the internet (*web*) were never significant in any of the eight models and had near-zero coefficients. The results for *fixed* support the results of many past studies (Allen, Carletti, et al., 2012; T. Beck et al., 2007; Sarma & Pais, 2010b), but the results for *mobile* and *web* contradict those same studies, as well as the result of other studies focused specifically on digital technology's effects on financial inclusion (SSonko, 2010). However this may lend evidence to the work of Kempson et al., which showed that while financial inclusion is positively impacted by growing digital access to financial services, it is also negatively impacted as financial institutions invest less in traditional, physical points of access to services (2004). The insignificant, near-zero coefficients for *mobile* and *web* may reflect this mixed effect of growing digital financial services.

The coefficients for the rural population of country (*rpop*) and the GINI coefficient (*gini*) – a measure of income inequality – do not follow expectations stated in 3.4.3. Both variables were expected to have negative relationships with financial inclusion, but both had weakly positive coefficients. Rurality is a proxy for distance from access points to financial services, and has been show in some cross sectional and country studies to be negatively correlated with financial inclusion (Carletti et al., 2012; Johnson & Arnold, 2012; Katoroogo, 2016). Though the models' results contradict these studies' conclusions, they may support the smaller body of work that argues that rurality and financial inclusion are, at most, unrelated (Chikalipah, 2017; Wentzel et al., 2013). Income inequality, on the other hand, is unambiguously shown to be negatively related to financial inclusion – which conflicts with these models' weakly positive coefficients for *gini* (Sarma & Pais, 2010b).

4.4.2 Accounting for Multicollinearity

As noted in section 4.3, some of the control variables are collinear. Collinearity in the regressors of a regression model can increase the variance of regression parameters – such as the coefficients that are the centre of this study's focus (Greene, 2003; Mela, 2002). It can also cause models to return inflated R^2 values despite low parameter significance and flip the signs of parameters in unexpected ways (Mela, 2002). The next step of the analysis accounts for multicollinearity between *fixed*, *web* and *lit* using a stepwise analysis. Note that both *lit* and *fixed* are collinear with *web* but are not collinear with each other. First, four models are built for each dependent financial inclusion variable - four for *BANK*, and four for *IFI* - which include *lit* and *fixed* but exclude *web*. Then four more models are built for *BANK* and *IFI* which include *web* but exclude *fixed* and *lit*. The resulting eight models for *BANK* are shown in Table 4.5 and the eight models for *IFI* are shown in Table 4.6.

The models of *BANK* change slightly after accounting for multicollinearity. Models that exclude *web* are nearly identical in terms of their diagnostic statistics, their R^2, and the signs and significance of the regressor coefficients. Three of the four models still have a significant ($p < 5\%$) linguistic diversity regressor (*LLG*, *ALE*, and *FRN*). The model including *LLG* now has a BP statistic that is significant at 10%, suggesting that the model's error terms exhibit heteroskedasticity. The model including *LPC* now has a significant ($p < 10\%$) coefficient for *rpop*, suggesting that in the absence of a variable representing internet connectivity like *web*, rurality becomes a more significant determinant of financial inclusion. The effects on the models of *BANK* which include *web* but exclude *lit* and *fixed* were more profoundly affected.

***Table 4.5** – Eight regression models for* BANK *accounting for multicollinearity of* lit *and* fixed *with* web

	Models of *BANK* using *lit* and *fixed* but not *web*								**Models of *BANK* using *web* but not *lit* and *fixed***							
	Coeff.	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*
INTERCEPT	(0.25) 0.32	(0.80)	(0.21) 0.34	(0.63)	(0.26) 0.31	(0.84)	(0.25) 0.33	(0.75)	0.19 0.20	0.33	0.16 0.20	0.80	0.17 0.20	0.89	0.21 0.20	1.04
LLG	0.21** 0.10	2.16	-	-	-	-	-	-	0.03 0.09	0.38	-	-	-	-	-	-
LPC	-	-	0.30 0.31	0.95	-	-	-	-	-	-	0.05 0.30	0.18	-	-	-	-
ALE	-	-	-	-	0.19** 0.08	2.50	-	-	-	-	-	-	0.08 0.07	1.20	-	-
FRN	-	-	-	-	-	-	0.22** 0.10	2.30	-	-	-	-	-	-	0.11 0.09	0.21
GDPPC	0.98*** 0.30	3.22	1.04*** 0.32	3.29	1.00*** 0.30	3.35	0.99*** 0.30	3.28	0.52*** 0.10	5.18	0.53*** 0.10	5.22	0.52*** 0.10	5.22	0.50*** 0.10	5.02
LIT	0.44 0.27	1.63	0.42 0.29	1.45	0.40 0.26	1.52	0.41 0.28	1.48	-	-	-	-	-	-	-	-
RPOP	0.19 0.13	1.54	0.23* 0.13	1.74	0.18 0.12	1.45	0.17 0.13	1.36	0.14 0.12	1.18	0.17 0.12	1.40	0.13 0.12	1.11	0.13 0.12	1.05
GINI	0.23 0.29	0.80	0.15 0.31	0.48	0.24 0.29	0.82	0.25 0.30	0.81	(0.20) 0.26	(0.75)	(0.16) 0.27	(0.58)	(0.20) 0.26	(0.76)	(0.20) 0.26	(0.76)
FIXED	0.58** 0.26	2.25	0.46 0.28	1.61	0.58* 0.25	2.31	0.52* 0.27	1.89	-	-	-	-	-	-	-	-
MOBILE	0.10 0.16	0.63	0.14 0.16	0.87	0.14 0.15	0.91	0.10 0.16	0.63	0.31** 0.14	2.30	0.31** 0.13	2.30	0.31** 0.13	2.35	0.26* 0.13	1.95
WEB	-	-	-	-	-	-	-	-	0.35** 0.13	2.63	0.36*** 0.14	2.67	0.36*** 0.13	2.73	0.36*** 0.13	2.77
NPL	(0.24) 0.51	(0.47)	(0.01) 0.52	(0.02)	(0.28) 0.51	(0.55)	(0.05) 0.50	(0.09)	(0.06) 0.43	(0.15)	(0.04) 0.43	(0.08)	(0.15) 0.42	(0.34)	(0.16) 0.41	(0.39)
R^2	0.60		0.55		0.61		0.59		0.67		0.64		0.64		0.68	
N	63		62		63		61		87		86		87		85	
	Test stat	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*
JB Test	1.72	0.42	0.57	0.75	1.54	0.46	1.67	0.43	1.10	0.58	0.90	0.64	1.38	0.50	1.20	0.55
BP Test	13.9*	0.08	9.28	0.32	10.91	0.21	9.14	0.33	16.6***	0.02	13.8*	0.05	13.4*	0.06	16.3**	0.02
F Test	10.1***	0.00	7.97***	0.00	10.5***	0.00	8.97***	0.00	22.6***	0.00	22.2**	0.00	23.2***	0.00	23.0***	0.00

Note: LLG = Largest Language Group; LPC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; gdppc= GDP per capita; lit = Adult literacy %; rpop = Rural population %; gini = Gini Coefficient; fixed = Fixed telephones per 100 inhabitants; mobile = Mobile telephones per 100 inhabitants; web = % population using internet; npl = Non-Performing Loan %; JB = Jarque-Bera Test; BP = Breusch-Pagan Test; Calculated with Python, author's elaboration; (Seabold & Perktold, 2010). **** denotes significance at p = 1%, ** at p = 5%, * at p = 10%;*

While the changes rendered all the models' linguistic diversity coefficients insignificant and nearly zero, it also led to heteroskedasticity in the models' errors terms, as evidenced by the significant BP statistics. For this reason, these final models were not used in the next step of the analysis. Rather, the models of *BANK* including *lit* and *fixed* and excluding *web* were used. Despite these models' slightly lower R^2, they comply with all the assumptions of regression analysis – including evasion of multicollinearity – and have high overall significance across many regressors. The exception is the model of *BANK* and *LLG*, which is either multicollinear in the initial model, or heteroskedastic in the latter stepwise model. Most of these models have positive, significant coefficients for the linguistic diversity variables – suggesting that the null hypothesis of this study – that linguistic diversity and financial inclusion are not related –may be rejected. However, this significant positive relationship does *not* appear in the best BANK-LPC model, calling this conclusion into doubt.

Another factor that supports the null hypothesis is the fact that a clear relationship between linguistic diversity and financial inclusion also does *not* appear in the stepwise models generated for the *IFI* dependent variable. Two of the four models of *IFI* with *lit* and *fixed* have linguistic diversity variables with significant coefficients ($p < 10\%$) – though both have marginal effects on the independent variable. All four of these models retain their good diagnostics, with no evidence to reject the assumptions of heteroskedasticity or Normality of the error terms. The R^2 of the models is unaffected, and these models of *IFI* do appear to be better specified than the initial models overall.

The last four models of *IFI*, in which *web* is included but *fixed* and *lit* excluded, do not have any significant linguistic diversity coefficients. However, two of the models have significant BP statistics, and so exhibit heteroskedasticity. The models have slightly higher R^2 terms than any of the previous models of *IFI*. The two models without heteroskedasticity (*IFI-LPC* and *IFI-FRN*) are thus the best specified models of these financial inclusion – linguistic diversity pairs, while the *IFI-LLG* and *IFI-ALE* models that include *lit* and *fixed* but exclude *web* appear to be the best specified for those pairs.

Table 4.6 *– Eight regression models for* IFI *accounting for multicollinearity of* lit *and* fixed *with* web

	Models of *IFI* using *lit* and *fixed* but not *web*								**Models of *IFI* using *web* but not *lit* and *fixed***							
	Coeff.	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*
INTERCEPT	(0.30)** 0.14	(2.14)	(0.25)* 0.15	(1.71)	(0.30)** 0.15	(2.07)	(0.26)* 0.15	(1.78)	0.05 0.11	0.46	0.03 0.11	0.28	0.05 0.11	0.46	0.09 0.11	0.83
LLG	0.07* 0.04	1.70	-	-	-	-	-	-	0.00 0.05	0.03	-	-	-	-	-	-
LPC	-	-	(0.07) 0.14	(0.48)	-	-	-	-	-	-	(0.22) 0.16	(1.38)	-	-	-	-
ALE	-	-	-	-	0.04 0.04	1.20	-	-	-	-	-	-	0.00 0.04	(0.08)	-	-
FRN	-	-	-	-	-	-	0.08* 0.04	1.79	-	-	-	-	-	-	0.02 0.05	0.34
GDPPC	0.46*** 0.14	3.35	0.48*** 0.14	3.42	0.47*** 0.14	3.42	0.46*** 0.14	3.41	0.32*** 0.06	5.88	0.31*** 0.05	5.79	0.32*** 0.06	5.88	0.30*** 0.06	5.50
LIT	0.40*** 0.12	3.25	0.34** 0.13	2.62	0.38*** 0.12	3.10	0.35*** 0.12	2.85	-	-	-	-	-	-	-	-
RPOP	0.10* 0.06	1.69	0.10* 0.06	1.75	0.10* 0.06	1.68	0.08 0.06	1.35	0.07 0.07	1.03	0.07 0.07	1.08	0.07 0.07	1.04	0.04 0.07	0.66
GINI	0.15 0.13	1.15	0.14 0.14	1.05	0.15 0.13	1.09	0.14 0.14	1.01	0.00 0.14	(0.01)	0.03 0.15	0.23	0.00 0.14	(0.01)	(0.04) 0.14	(0.31)
FIXED	0.24* 0.12	2.04	0.23* 0.12	1.87	0.22* 0.12	0.188	0.25** 0.12	2.04	-	-	-	-	-	-	-	-
MOBILE	0.09 0.07	1.22	0.12 0.07	1.62	0.10 0.07	1.47	0.08 0.07	1.16	0.15** 0.07	2.09	0.16** 0.07	2.23	0.15** 0.07	2.12	0.13* 0.07	1.81
WEB	-	-	-	-	-	-	-	-	0.22*** 0.07	3.01	0.23*** 0.07	3.13	0.22*** 0.07	2.99	0.21*** 0.07	2.95
NPL	(0.25) 0.23	(1.09)	(0.14) 0.23	(0.61)	(0.22) 0.24	(0.94)	(0.21) 0.22	(0.95)	(0.36) 0.23	(1.57)	(0.28) 0.23	(1.21)	(0.36) 0.23	(1.55)	(0.42)* 0.23	(1.86)
R^2	0.70		0.68		0.69		0.71		0.73		0.74		0.73		0.73	
N	63		62		63		61		87		86		87		85	
	Test stat	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*	*Test stat*	*p-value*
JB TEST	1.89	0.39	0.85	0.65	1.10	0.58	0.33	0.85	1.87	0.39	1.45	0.48	1.91	0.39	1.87	0.39
BP TEST	12.99	0.11	8.65	0.37	8.97	0.34	7.48	0.49	15.2**	0.03	8.62	0.28	12.1*	0.09	11.5	0.12
F TEST	15.6***	0.00	14.3***	0.00	15.1***	0.00	15.6***	0.00	30.4***	0.00	30.9***	0.00	30.4***	0.00	30.3***	0.00

Note: LLG = Largest Language Group; LPC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; gdppc= GDP per capita; lit = Adult literacy %; rpop = Rural population %; gini = Gini Coefficient; fixed = Fixed telephones per 100 inhabitants; mobile = Mobile telephones per 100 inhabitants; web = % population using internet; npl = Non-Performing Loan %; JB = Jarque-Bera Test; BP = Breusch-Pagan Test; Calculated with Python, author's elaboration; (Seabold & Perktold, 2010). **** denotes significance at p = 1%, ** at p = 5%, * at p = 10%;*

4.4.3 Seemingly Unrelated Regression

The final step of this study's analysis is the construction of Seemingly Unrelated Regression (SUR) models, which combine the four best models for each dependent financial inclusion variable – one for each possible linguistic diversity variable – into a system of equations. The intent of this final step is to perform a joint re-estimation of the component OLS models, and to use the resulting models to test this study's hypothesis.

The models selected to be used in each SUR model are shown in Table 4.7. Note that two initial models, despite their multicollinearity, were used in the SUR 1 model of *BANK*. This was done only in cases when the models built in the stepwise approach did not generate a robust model for that pair of *BANK* – linguistic diversity variables, often due to heteroskedasticity. The reason the multicollinear models were preferred over heteroskedastic models is that joint estimation through the SUR technique can actually lead to great efficiency gains in cases where models are multicollinear (Binkley, 982; Binkley & Nelson, 1988; Zellner & Lee, 1962).

Table 4.7 *– Component OLS Model Selection for SUR Models*

		OLS Model		
		Initial Model	***Fixed & Lit* – No *Web* Model**	***Web* – No *Fixed & Lit* Model**
SUR 1 (*BANK*)	***BANK – LLG***	YES	Heteroskedasticity	Heteroskedasticity
	BANK – LPC	Multicollinearity	YES	Heteroskedasticity
	BANK – ALE	Multicollinearity	YES	Heteroskedasticity
	BANK – FRN	YES	Heteroskedasticity	Heteroskedasticity
SUR 2 (*IFI*)	***IFI – LPC***	Multicollinearity	YES	Heteroskedasticity
	IFI – LPC	Multicollinearity	Less R^2 than alternative	YES
	IFI – ALE	Multicollinearity	YES	Heteroskedasticity
	IFI – FRN	Multicollinearity	YES	Heteroskedasticity

Note: 'YES' indicates the model was used in the subsequent SUR analysis for either BANK *(SUR 1) or for* IFI *(SUR 2). Other cells give the reason why a model was not selected for inclusion in the SUR models.*

The results of the SUR estimations are displayed in Table 4.8. Despite having lower overall R^2 than the component models, the resulting models are better specified than the individual OLS models, with more significant regressors than the previous models.

Most interestingly, none of the linguistic diversity regressors in the SUR-adjusted models are significant, and their coefficients are all nearly zero. This supports the null hypothesis that linguistic diversity is not related to financial inclusion. The other regressors, however, now have larger, more significant coefficients. In the resulting models of *BANK*, coefficients of *gdppc*, *rpop*, and *fixed* are all positive and highly significant ($p < 1\%$). *Mobile* is also positive and significant at 5%. This is all consistent with the bulk of academic literature on the relationship between these variables and financial inclusion at a country level, as discussed in the previous section (4.4.2), as well as in section 3.3.4. However, the SUR-derived models are not perfectly aligned with past literature. Adult literacy, despite having positive coefficients across all four *BANK* models, is not significant in any of them, despite frequent identification as an important determinant in past studies that include the same group of variables as these models. The GINI coefficient is also positive and significant at 5%, which contradicts much of the literature on inequality's detrimental effect on financial inclusion (García-Herrero & Martínez Turégano, 2015; Sarma & Pais, 2010b). The sign of the variable for non-performing loans (*npl*) also switched after the SUR estimation, though its coefficients are not significant.

The models of *IFI* also changed after re-estimation in SUR 2. The linguistic diversity variables are also nearly zero and insignificant, which again supports this study's null hypothesis. It also speaks to the consistent absence of an apparent relationship between these two operationalizations of financial inclusion (*BANK* and *IFI*) and the four operationalizations of linguistic diversity (*LLG*, *LPC*, *ALE*, *FRN*). As with the *BANK* models from SUR 1, the coefficients of many of the control variables have significant, positive coefficients, as one would expect from the classical determinants of financial inclusion: *gdppc*, *lit*, *mobile*, and *web*. Unlike the *BANK* models, *npl* once again has the expected negative relationship with *IFI*, though the coefficients are again insignificant. Unexpectedly, *rpop* and *gini* also have significant, positive coefficients.

The important conclusion drawn from these eight component models from the two SUR re-estimations is that across all of them, the variables representing linguistic diversity appear to not be significant determinants of country-level financial inclusion – even when both those concepts are measured using different variables. This can be understood as a confirmation that the null hypothesis of this study cannot be rejected. While significant positive relationships appear between financial inclusion and linguistic diversity variables in some, unadjusted OLS models, the SUR re-estimated models evince a consistent absence of any relationship between these concepts.

Table 4.8 *– Seemingly Unrelated Regression Model Results for BANK and IFI*

	Results of SUR 1 for *BANK*								Results of SUR 2 for *IFI*							
	Coeff.	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*	*Coeff.*	*t*
INTERCEPT	(0.48) 0.29	(1.67)	(0.49) 0.29	(1.66)	(0.48)* 0.28	(1.72)	(0.48) 0.28	(1.67)	(0.26)*** 0.08	(3.09)	(0.04) 0.06	(0.70)	(0.27)*** 0.08	(3.21)	(0.26)*** 0.08	(3.09)
LLG	**0.01** **0.03**	**0.35**	-	-	-	-	-	-	**0.01** **0.01**	**0.53**	-	-	-	-	-	-
LPC	-	-	**0.00** **0.08**	**0.06**	-	-	-	-	-	-	**(0.08)** **0.08**	**(1.10)**	-	-	-	-
ALE	-	-	-	-	**0.01** **0.02**	**0.56**	-	-	-	-	-	-	**0.00** **0.01**	**0.43**	-	-
FRN	-	-	-	-	-	-	**0.01** **0.03**	**0.44**	-	-	-	-	-	-	**0.01** **0.01**	**0.53**
GDPPC	1.13*** 0.31	3.70	1.14*** 0.31	3.64	1.13*** 0.30	3.78	1.13*** 0.30	3.73	0.48*** 0.07	6.55	0.37*** 0.03	12.9	0.49*** 0.07	6.56	0.48*** 0.07	6.55
LIT	0.27 0.25	1.08	0.26 025	1.03	0.26 0.24	1.11	0.26 0.25	1.07	0.20*** 0.07	3.04	-	-	0.20*** 0.07	3.01	0.20*** 0.07	3.04
RPOP	0.39*** 0.11	3.52	0.40*** 0.11	3.55	0.39*** 0.11	3.63	0.39*** 0.11	3.52	0.13*** 0.04	3.39	0.09** 0.04	2.11	0.13*** 0.04	3.52	0.13*** 0.04	3.39
GINI	0.59** 0.26	2.30	0.61** 0.26	2.32	0.60** 0.25	2.38	0.60** 0.26	2.32	0.21** 0.09	2.28	0.11 0.10	1.07	0.22** 0.09	2.37	0.21** 0.09	2.28
FIXED	0.66*** 0.20	3.34	0.67*** 0.21	3.23	0.67*** 0.19	3.44	0.66*** 0.20	3.27	0.09 0.05	1.60	-	-	0.08 0.05	1.56	0.09 0.05	1.60
MOBILE	0.34** 0.14	2.37	0.35** 0.15	2.39	0.34** 0.14	2.45	0.34** 0.14	2.39	0.32*** 0.05	6.52	0.32*** 0.05	6.17	0.33*** 0.05	6.69	0.32*** 0.05	6.52
WEB	0.00 0.05	(0.02)	-	-	-	-	0.00 0.05	0.02	-	-	0.14*** 0.04	3.39	-	-	-	-
NPL	0.32 0.49	0.65	0.34 0.50	0.68	0.31 0.48	0.65	0.32 0.49	0.67	(0.17) 0.16	(1.07)	(0.25) 0.15	(1.62)	(0.17) 0.16	(1.02)	(0.17) 0.16	(1.07)
$\mathbf{R^2}$	0.53		0.49		0.53		0.49		0.59		0.58		0.59		0.59	
N	63		62		63		61		61		86		63		61	

*Note: *** denotes significance at p = 1% ** at p = 5%, * at p = 10%; LLG = Largest Language Group; LPC = Languages per Capita; ALE = Alesina's Ethnolinguistic Fractionalization Index; FRN = Fearon's Linguistic Diversity Index; gdppc= GDP per capita; lit = Adult literacy %; rpop = Rural population %; gini = Gini Coefficient; fixed = Fixed telephones per 100 inhabitants; mobile = Mobile telephones per 100 inhabitants; web = % population using internet; npl = Non-Performing Loan %; Calculated with R, author's elaboration* (R Development Core Team, 2007)

Chapter 5: Conclusions and Recommendations

5.1 Introduction

The aim of the research was to examine the relationship between linguistic diversity and financial inclusion at a country level. This chapter presents the conclusions of the research, drawn from the findings of the tests carried out in Chapter 4 above. It also acknowledges limitations of this study and recommends future avenues of research for related topics.

5.2 Summary of the Study

This study examined the relationship between financial inclusion and linguistic diversity at a country level. The study was explanatory and quantitative in nature and aimed to answer the question of whether a country level relationship existed between these concepts. One important challenge in answering this question was the broadness with which "linguistic diversity" and "financial inclusion" could be defined, let alone operationalized as variables. To overcome this challenge, several different variables were operationalized for each concept, using data and techniques to approximate both concepts from past research. The study's null hypothesis was that no relationship would be observed between any of the various combinations of these different operationalizations of linguistic diversity and financial inclusion.

The study was motivated by a perceived gap in literature linking financial inclusion to language and linguistic diversity. There are some studies of the relationship between linguistic diversity and other socioeconomic phenomena, such as levels of economic development. There are even some studies of financial inclusion that include language-related variables in their analyses, such as home language or literacy. But there were no country-level studies of the relationship between these two concepts. Such knowledge would be useful in ongoing efforts to increase financial inclusion around the world, especially in less-developed, linguistically diverse polities where there is a *perceived* trade-off between linguistic diversity and economic "progress". It would be especially useful to balance and adapt efforts to both increase financial inclusion *and* preserve the cultural and social value of language and linguistic diversity.

The study tested this hypothesis by building OLS models between different combinations of dependent financial inclusion variables and independent linguistic diversity variables, controlling for known determinants of country level financial inclusion. Additional models were built to control for multicollinearity in some of the control variables. For each unique combination of a financial inclusion variable and a linguistic diversity variable, the best model was selected to become a component model in a seemingly unrelated regression model – one for each of the two financial inclusion variables. These selections were made based on whether certain models met the assumptions of OLS regression, the models' overall significance, and their R^2. The SUR models produced joint re-estimations of all the selected component models, accounting for correlation between the models' error terms. Overall, the results of the study suggest that there is indeed no discernible, consistent relationship between financial inclusion and linguistic diversity at a country level.

5.3 Conclusion

The study did not identify a consistent, significant relationship between any single pair of variables representing the concepts of financial inclusion and linguistic diversity, much less a consistent relationship across all the operationalized variables.

All the coefficients of correlation calculated in section 4.3 indicated that linguistic diversity had a weakly negative relationship with financial inclusion across the operationalized variables for both concepts. This would mirror the relationship of linguistic diversity to economic development as postulated by the Fishman-Pool hypothesis, which implies that linguistic diversity varies inversely with economic development (Fishman, 1966; Nettle, 2000; Pool, 1972). This would be plausible, given the known positive relationship between economic development and financial inclusion identified in literature on those concepts (Chikalipah, 2017; Demirgüç-Kunt et al., 2018).

Subsequent regression analyses rarely returned significant, explanatory coefficients for the linguistic diversity variables. When they did, these results seemed to contradict the results of the correlation analysis, since all the superficially significant relationships that emerged were weakly positive in direction. After the best OLS models were combined into SUR models, none of the resulting re-estimated models yielded a significant coefficient of regression linked to linguistic diversity. Furthermore, all those insignificant coefficients were within 0.1 of zero – indicating an almost complete absence of any explanatory relationship with the dependent financial inclusion variables. In conclusion, there was not enough evidence to reject the null hypothesis that no relationship exists between linguistic diversity and financial inclusion.

5.4 Limitations

This study faced serious limitations for the perspective of data availability.

- Language data is especially difficult to collect at a country-level across a consistent time period, which renders panel studies nearly impossible, and raises methodological concerns for simpler cross-sectional studies.
- There were limitations on the availability and timeliness of data for many of the control variables. This, in addition to the step-wise approach taken to eliminate multicollinearity of regressors in the models, led to models with different, reduced sample sizes.

It also faced limitations in that, as discussed in Chapters 2 and 3, conceptualizing and operationalizing financial inclusion and linguistic diversity is difficult – due both to the data availability problem mentioned, but also due to the broadness and multifaceted nature of both concepts. This study has tried to create varied, robust operationalizations of both concepts to account for this, but other options for operationalizing either concept exist as well. The data limitations also meant that the size of the sample of countries included in the study was fairly small, which may limit the general applicability of the results to all countries.

Another limitation is that the control variables included in this study are not exhaustive. There may be other, confounding variables that should have been included in the study as controls.

5.5 Recommendations

Evidence for the *absence* of a relationship is not very actionable from a business or policy standpoint. One potential recommendation would be that any business – such as a financial institution operating in a linguistically diverse context – or government agency – such as one responsible for language policy in commerce or education – should not consider linguistic diversity a natural enemy of financial inclusion, as there is not sufficient evidence that the concepts are negatively related.

However, the results of this study are very actionable from an academic standpoint – there are many potential avenues for future research, especially if more complete language data can be gathered. Time series data regarding adoption of languages in countries, or changing language mixes in populations, would allow this study to be repeated using panel data rather than simple cross-sectional data. In addition, future research could examine the role of language in more specific types of financial inclusion, such as digital financial services, or in an aspect of financial inclusion, such as access to credit. While a relationship to financial inclusion as an entire concept does not appear to exist, one may be identified between aspects of financial inclusion that are mediated through written, official language, such as mobile money applications or loan contracts.

Perhaps the most interesting avenue of research is represented by the unique case studies cited in section 2.6.4, which details how some financial institutions turned linguistic diversity into an opportunity to engage and retain minority language speaking customers, diversify their balance sheets, and gain an advantage over their competitors. Identifying common elements of these successes that may be replicated could be key to catalysing financial inclusion in other linguistically diverse areas, while helping to preserve that same diversity.

Works Referenced

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, *91*(5), 1369–1401. https://doi.org/10.1257/aer.91.5.1369

Agency, C. I. (2019). *The World Factbook*. The World Factbook 2019. https://www.cia.gov/library/publications/resources/the-world-factbook/index.html

Akudugu, M. . (2013). The Determinants of Financial Inclusion in Western Africa : Insights from Ghana. *Research Journal of Finance and Accounting ISSN 2222-2847*, *4*(8), 1–10.

Alesina, A., Devleeschauwer, A., Easterly, W., & Kurlat, S. (2003). *Fractionalization*. 155–194.

Allen, F., Carletti, E., Cull, R., Qian, J., Senbet, L., & Valenzuela, P. (2012). Resolving the African Financial Development Gap: Cross-Country Comparisons and a Within-Country Study of Kenya. In S. Edwards, S. Johnson, D. Weil, & University of Chicago Press (Eds.), *NBER African Successes 2011* (pp. 13–62). National Bureau of Economic Research. https://doi.org/10.7208/chicago/9780226315867.003.0003

Allen, F., Demirgüç-Kunt, A., Klapper, L., & Martinez-Peria, M. S. (2012). The Foundations of Financial Inclusion: Understanding Ownership and Use of Formal Accounts. World Bank Policy Research Paper No. 6290. In *World Bank Policy Research Working Paper* (Vol. 27, Issue December).

Amaeshi, K. M. (2011). Financial Exclusion, Financial Institutions and Corporate Social Responsibility: A Developing Country Perspective. *SSRN Electronic Journal*, 1–15. https://doi.org/10.2139/ssrn.950989

Arcand, J., & Grin, F. (2013). Language in economic development : Is English special and is linguistic fragmentation bad ? *English and Development. Policy, Pedagogy and Globalization*, 243–266. https://archive-ouverte.unige.ch/unige:40075

Arcand, J. L. (1996). Development economics and language: The earnest search for a mirage? *International Journal of the Sociology of Language*, *121*(1), 119–158. https://doi.org/10.1515/ijsl.1996.121.119

Asfaw, H. A. (2015). Financial Inclusion through Mobile Banking: Challenges and Prospects. *International Journal of Scientific Research in Science and Technology*, *4*(1), 232–238. www.ijsrst.com

Atkinson, A. (2006). *Migrants and Financial Services : A review of the situation in the United Kingdom* (Issue March). https://www.bristol.ac.uk/medialibrary/sites/geography/migrated/documents/pfrc0605.pdf

Atkinson, A., & Messy, F.-A. (2013). Promoting Financial Inclusion through Financial Education: OECD/INFE Evidence, Policies and Practice. *OECD Publishing*, *34*(34), 1–53. https://doi.org/10.1787/5k3xz6m88smp-en

Barcellos, S. H., Smith, J. P., Yoong, J. K., & Carvalho, L. (2012). *Barriers to Immigrant Use of Financial Services: The Role of Language Skills, US Experience, and Return Migration Expectations* (WR-923-SSA).

Beck, T., Demirgüç-Kunt, A., Soledad, M., & Peria, M. (2007). Reaching Out: Access To And Use of Banking Services Across Countries. *Journal of Financial Economics*, *85*(1), 234–266.

Beck, Thorsten, Demirgüç-Kunt, A., & Peria, M. S. M. (2008). Banking services for everyone? Barriers to bank access and use around the world. *World Bank Economic Review*, *22*(3), 397–430. https://doi.org/10.1093/wber/lhn020

Bihari, S. C. (2011). Financial Inclusion for Indian Scene. *SCMS Journal of Indian Management*. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=3934286&site=ehost-live

Binkley, J. K. (982). The effect of variable correlation on the efficiency of seemingly unrelated regression in a two-equation model. *Journal of the American Statistical Association1*, *77*, 890–895.

Binkley, J. K., & Nelson, C. H. (1988). A note on the efficiency of seemingly unrelated regression. *American Statistician*, *42*, 137–139.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287. https://doi.org/10.2307/1911963

Brock-Utne, B. (2001). Education for All: In Whose Language? *Oxford Review of Education*, *27*(1), 115–134. https://doi.org/10.1080/3054980020030637

Cámara, N., & Tuesta, D. (2017). *Measuring financial inclusion: a multidimensional index 1 Measuring financial inclusion: a multidimensional index*. *July*.

Carletti, E., Senbet, L. W., Cull, R., Allen, F., Qian, J., & Valenzuela, P. (2012). Improving Access to Banking: Evidence from Kenya. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3305047

Casson, R. J., & Farmer, L. D. M. (2014). Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clinical and Experimental Ophthalmology*, *42*(6), 590–596. https://doi.org/10.1111/ceo.12358

Chaibou, O. I. (2019a). *Rôle de la microfinance dans l'inclusion financière des pays de l'UEMOA : application au cas du Niger*. Universitéde Roeun Normandie.

Chaibou, O. I. (2019b). *Rôle de la microfinance dans l'inclusion financière des pays de l'UEMOA : application au cas du Niger*.

Chen, F. W., Feng, Y., & Wang, W. (2018). Impacts of financial inclusion on non-performing loans of commercial banks: Evidence from China. *Sustainability (Switzerland)*, *10*(9), 1–28. https://doi.org/10.3390/su10093084

Chibba, M. (2009). Lessons from selected development policies and practices. *Development in Practice*, *19*(3), 365–370. https://doi.org/10.1080/09614520902808100

Chikalipah, S. (2017). What determines financial inclusion in Sub-Saharan Africa? *African Journal of Economic and Management Studies*, *8*(1), 8–18. https://doi.org/10.1108/AJEMS-01-2016-0007

Conroy, J. (2005). APEC and Financial Exclusion: Missed Opportunities for Collective Action? *Asia-Pacific Development Journal*, *12*(1), 28–29.

Coyne, G. (2015). Language Education Policies and Inequality in Africa: Cross-National Empirical Evidence. *Comparative Education Review*, *59*(4), 619–637. https://www.jstor.org/stable/pdf/1050997.pdf?refreqid=excelsior%3Af5a923f6ea0e7a4d182b2234d64656b9

Dabla-Norris, E., Ji, Y., Townsend, R. M., & Unsal, D. F. (2015a). *Distinguishing constraints on financial inclusion and their impact on GDP, TFP, and inequality* (No. 20821).

Dabla-Norris, E., Ji, Y., Townsend, R., & Unsal, D. F. (2015b). *Identifying constraints to financial inclusion and their impact on GDP and inequality: a structural framework for policy* (WP/15/22). https://doi.org/10.3386/w20821

Demirgüç-Kunt, A., & Klapper, L. (2013). Measuring Financial Inclusion: Explaining Variation in Use of Financial Services across and within Countries. *Brookings Papers on Economic Activity*, *2013*(1), 279–340. https://doi.org/10.1353/eca.2013.0002

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. In *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. https://doi.org/10.1596/978-1-4648-1259-0

Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. https://doi.org/10.1596/978-1-4648-0552-3

Desmet, K., Ortuno-Ortín, I., & Wacziarg, R. (2016). Linguistic cleavages and economic development. *The Palgrave Handbook of Economics and Language*, 425–446. https://doi.org/10.1007/978-1-137-32505-1

Desmet, K., Ortuno-ortín, I., & Weber, S. (2009). Linguistic Diversity and Redistribution. *Oxford University Press, 7*(6), 1291–1318. http://www.jstor.org/stable/40601204

Escalante, J. D. R., & Batun, R. Y. (2018). *La inclusión financiera maya*. Universidad de Quintana Roo.

Evans, O., & Adeoye, B. (2016). Determinants of Financial Inclusion in Africa : A Dynamic Panel Data Approach. *University of Mauritius Research Journal*, *22*(81326).

Eves, R., & Titus, A. (2017). *Gender Challenges to Financial Inclusion in Papua New Guinea*. http://ssgm.bellschool.anu.edu.au/sites/default/files/publications/attachments/2017-12/dp-2017-8_evestitus_online_rev1.pdf

Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, *8*(2), 195–222.

Fiebig, D. G. (2003). Seemingly Unrelated Regression. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 101–121). Blackwell Publishing. http://www.uomisan.edu.iq/admineco/ar/admin/pdf/5882582627.pdf#page=119

Finau, G., Rika, N., Samuwai, J., & McGoon, J. (2016). Perceptions of Digital Financial Services in Rural Fiji. *Information Technologies and International Development*, *12*(4), 11–21.

Fishman, J. (1964). Some Contrasts Between Linguistically Homogeneous and Linguistically Heterogeneous Polities *. *Sociological Inquiry*, 146–159.

Fishman, J. (1966). Some contrasts between linguistically homogeneous and linguistically hetergeneous polities. *Sociological Inquiry*, *36*(2), 146–158.

Furusawa, M. (2016). Can Financial Inclusion Make a Difference in Central America? *IMF Central America Regional Conference*. https://www.imf.org/en/News/Articles/2016/11/17/SP111716-Can-Financial-Inclusion-Make-a-Difference-in-Central-America

García-Herrero, A., & Martínez Turégano, D. (2015). *Financial inclusion, rather than size, is the key to tackling income inequality* (15/05). https://www.bbvaresearch.com/wp-content/uploads/2015/02/WP_Financial-Inclusion-Income-Inequality4.pdf

Gibbs, J. (2010). *Financial Inclusion amongst New Migrants in Northern Ireland*.

Global Partnership for Financial Inclusion. (2012). *G20 Financial Inclusion Indicators*. http://www.gpfi.org/sites/default/files/G20 Set of Financial Inclusion Indicators.pdf

Goedecke, J., Guérin, I., D’Espallier, B., & Venkatasubramanian, G. (2018). Why do financial inclusion policies fail in mobilizing savings from the poor? Lessons from rural south India. *Development Policy Review, 36,* O201–O219. https://doi.org/10.1111/dpr.12272

Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the Behavioral Sciences Student Edition* (9th ed.). Cengage Learning.

Greenberg, J. H. (1956). The Measurement of Linguistic Diversity. *Language*, *32*(1), 109–115.

Greene, W. H. (2003). *Econometric Analysis* (5th ed.). Prentice-Hall.

Grin, F. (2007). Economics and language policy. In *Handbook of Language and Communication - Diversity and Change* (pp. 271–297). Mouton de Gruyter.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Hall, R. E., & Lilien, D. M. (2014). *EViews 8.1 User Guide I & II* (8.1). IHS Global Inc. http://www.eviews.com/EViews8/EViews8/EViews 8 Users Guide I.pdf

Hammarstrom, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution*, 19–29. https://doi.org/10.1093/jole/lzw002

Herfindahl, O. (1950). *Concentration in the U.S. Steel Industry*. Columbia University.

Honohan, P., & King, M. (2009). Case and Effect of Financial Access: Cross-Country Evidence from the Finscope Surveys. *Measurement, Promotion, and Impact of Access to Financial Services*.

Hutchinson, W. K. (2005). “Linguistic Distance” as a Determinant of Bilateral Trade. *Southern Economic Journal*, *72*(1), 1. https://doi.org/10.2307/20062091

Isphording, I. E., & Otten, S. (2013). The costs of babylon-linguistic distance in applied economics. *Review of International Economics*, *21*(2), 354–369. https://doi.org/10.1111/roie.12041

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo Evidence. *Economics Letters*, *6*(3), 255–259.

Johnson, S., & Arnold, S. (2012). Inclusive financial markets: Is transformation under way in Kenya? In *Bath Papers in International Development and Wellbeing* (No. 14; Vol. 14). http://hdl.handle.net/10419/128109

Jossefa, A. L. (2011). *Determinantes do Acesso ao Sistema Financeiro: o Caso do Moçambique*. Universidade Técnica de Lisboa.

Katoroogo, R. M. (2016). *Behavioural determinants of financial inclusion in Uganda* [University of the Witswatersrand]. https://pdfs.semanticscholar.org/fcaf/92c40053eaedbfe574eeabb9709e07c7406d.pdf?_ga=2.260355917.1398208648.1594010013-173367550.1594010013

Kempson, E., Atkinson, A., & Pilley, O. (2004). Policy level response to financial exclusion in developed economies : lessons for developing countries. In *Personal Finance* (Issue September).

Kempson, E., & Whyley, C. (1999). *Understanding and combating financial exclusion* (Issue March).

Kennedy, P. (2008). *A Guide to Econometrics* (6th ed.). Blackwell Publishing.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Prentice-Hall.

Kosonen, K. (2017). *Language of instruction in Southeast Asia*.

Kouame, K. (2014). *La stratégie marketing de communication commerciale dans le cadre de l'activité de la microfinance ivoirienne : un démenti à l'exoglossie ? 2000*, 26–28.

Kwok, A. (2019, December). Lost in Tongue: The politics of mother tongue education in Myanmar. *New Mandala*. https://www.newmandala.org/lost-in-tongue/

Laitin, D., & Ramachandran, R. (2014). *Language Policy and Economic Development*. *94305*(March).

Lee, E., & Chang, A. (2019). *The Challenge of Teaching Pacific Islanders to Save*. Pacific Financial Inclusion Program Blog. http://www.pfip.org/newsroom/blog/the-challenge-of-teaching-pacific-islanders-to-save/

Leyshon, A., & Thrift, N. (1995a). Geographies of Financial Exclusion: Financial Abandonment in Britain and the United States. *Transactions of the Institute of British Geographers*, *20*(3), 312. https://doi.org/10.2307/622654

Leyshon, A., & Thrift, N. (1995b). *Geographies of Financial Exclusion: Financial Abandonment in in Britain and the United States*. *20*(3), 312–341.

Lian, B., & Oneal, J. R. (1997). Cultural Diversity and Economic Development: A Cross-National Study of 98 Countries, 1960-1985. *Economic Development and Cultural Change*, *46*(1), 61–78. https://doi.org/10.2307/259249

Loh, J., & Harmon, D. (2010). *The Index of Linguistic Diversity : A New Quantitative Measure of Trends in the Status of the World ' s Languages* (Vol. 4).

Lutheran Immigration and Refugee Service, & RefugeeWorks. (2005). *Financial literacy for newcomers: weaving immigrant needs into financial education*. https://citiesofmigration.ca/elibrary/financial-literacy-for-newcomers-weaving-immigrant-needs-into-financial-education-2/

McDonnell, S. (2003). Money Talks: Overcoming the financial exclusion problems faced by Indigenous Australians. *Australian Social Policy Conference, July*.

Mela, C. F. (2002). The Impact of Collinearity on Regression Analysis : The Asymmetric Effect of The impact of collinearity on regression analysis : the asymmetric e V ect of negative and positive correlations. *Applied Economics*, *March 2002*. https://doi.org/10.1080/00036840110058482

Mohan, R. (2006). *Economic growth, financial deepening, and financial inclusion*.

Morgan, P., & Pontines, V. (2014). Financial Stability and Financial Inclusion. *SSRN Electronic Journal*, *488*. https://doi.org/10.2139/ssrn.2464018

Mramba, N., Sutinen, E., Haula, M., & Msami, P. (2014). Survey of mobile phone usage patterns among street vendors in Dar Es Salaam city - Tanzania. *International Journal of Information Technology and Business Management*, *28*(1), 1–10.

Naghavi, N. (2019). *State of the Industry Report on Mobile Money*. https://doi.org/10.1002/9781118290743.wbiedcs023

Nau, R. (2014). *Notes on Linear Regression*. Notes.

Nettle, D. (2000). Linguistic Fragmentation and the Wealth of Nations: The Fishman-Pool Hypothesis Reexamined. *Economic Development and Cultural Change*, *48*(2), 335–348. https://doi.org/10.1086/452461

Nettle, D., Grace, J. B., Choisy, M., Cornell, H. V., Guégan, J. F., & Hochberg, M. E. (2007). Cultural diversity, economic development and societal instability. *PLoS ONE*, *2*(9), 5–9. https://doi.org/10.1371/journal.pone.0000929

Oikocredit. (2019). *Banrural, Banco de Desarrollo Rural SA*. Project Partner Overview. https://www.oikocredit.coop/products-services/equity-investments/equity-partner/11137/banrural-banco-de-desarrollo-rural-sa

Otieno, O. C., Liyala, S., Odongo, B. C., & Abeka, S. O. (2016). Challenges Facing the Use and Adoption of Mobile Phone Money Services. *World Journal of Computer Application and Technology*, *4*(1), 8–14. https://doi.org/10.13189/wjcat.2016.040102

Ottaviano, G. I. P., & Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, *6*(1), 9–44. https://doi.org/10.1093/jeg/lbi002

Pleite, F. M.-C., Soriano, A. M. M., Eid, M., & Sueiras, J. C. (2016). Inclusión financiera en el ámbito rural mediante cajas de ahorro: Estudio de una experiencia en México. *Perfiles Latinoamericanos*, *24*(48). http://www.scielo.org.mx/scielo.php?pid=S0188-

76532016000200185&script=sci_arttext

Pool, J. (1972). National development and language diversity. *Advances in the Sociology of Language*, *II*, 213–230.

Posner, D. N. (2004). Measuring ethnic fractionalization in Africa. *American Journal of Political Science*, *48*(4), 849–863. https://doi.org/10.1111/j.0092-5853.2004.00105.x

R Development Core Team. (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rumanyika, J. D. (2015). *Obstacles Towards Adoption of Mobile Banking in Tanzania: a Review*. *35*(1), 1–17.

Salazar-Cantú, J., Jaramillo-Garza, J., & Rosa, B. Á.-D. la. (2015). Financial Inclusion and Income Inequality in Mexican Municipalities. *Open Journal of Social Sciences*, *03*(12), 29–43. https://doi.org/10.4236/jss.2015.312004

Sarma, M. (2008). *Index of Financial Inclusion* (No. 215). https://www.econstor.eu/handle/10419/176233

Sarma, M., & Pais, J. (2010a). Financial Inclusion and Development: A Cross Country Analysis. *In Annual Conference of the Human Development and Capability Association, New Delhi*, *10–13*, 1–30. https://doi.org/10.1002/jid

Sarma, M., & Pais, J. (2010b). Financial Inclusion and Development. *Journal of International Development*, *23*, 613–628. https://doi.org/10.1002/jid

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*.

Simons, G. F., & Fennig, C. D. (2018). *Ethnologue: Languages of the World, Twenty-first edition*. Ethnologue. http://www.ethnologue.com

Singh, C., Ram, K., Chandrakar, R., Mittal, A., Goenka, A., Goud, C. R. P., Suresh, R. V., Garg, R., & Kumar, U. (2014). *Financial Inclusion in India: Select Issues* (No. 474). https://www.iimb.ac.in/sites/default/files/2018-07/WP_No._474.pdf

Skutnabb-Kangas, T., Várady, T., Philipson, R., & Kontra, M. (1999). Linguistic diversity, human rights and the “free” market. *Kontra*, 187–222.

SSonko, G. W. (2010). *The Role of Mobile Money Services in Enhancing Financial Inclusion in Uganda*. *08/10*, 1–43.

The International Monetary Fund. (2018). *Financial Access Survey*. The International Monetary Fund\. data.imf.org/FAS

The pandas development team. (2020). *pandas-dev/pandas: Pandas 1.0.3 (Version v1.0.3)*. Zenodo. http://doi.org/10.5281/zenodo.3715232

The World Bank Group. (2016a). *Enterprise Surveys*. The World Bank Group. https://www.enterprisesurveys.org/data

The World Bank Group. (2016b). *Global Payment Systems Survey*. The World Bank Group. http://www.worldbank.org/en/topic/financialinclusion/brief/gpss

Trivelli A, C., & Tarazona S, Á. (2007). *Riesgos y Portafolios Agropecuarios: Lecciones desde la experiencia de instituciones financieras de América Latina* (No. 151). http://161.132.112.85/bitstream/IEP/779/2/documentodetrabajo151.pdf

Ulwodi, D. W. (2017). *Financial Inclusion and Welfare: Evidence From the Global Findex and Financial Access Surveys*. 1–188.

UNESCO. (2006). *Global Linguistic Diversity*. UNESCO Web Page. http://www.unesco.org/new/en/communication-and-information/resources/news-and-in-focus-articles/in-focus-articles/2006/measuring-linguistic-diversity-on-the-internet/global-linguistic-diversity/

UNESCO. (2016). *If you don't understand, how can you learn?* (No. 24; Global Education Monitoring Report, Issue February).

United Nations Statistics Division. (2018). *Demographic Yearbook*. United Nations. https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml#censusdatasets

United States Government Accountability Office. (2010). Consumer Factors Affecting the Financial Literacy of Individuals with Limited English Proficiency. In *USGAO Report to Congressional Committees* (Issue May).

Vicente, K. N. de S. (2011). *Análise da viabilidade de implementação e desenvolvimento de Microsseguros de Saúde em Angola (município da Matala)*. 74. http://hdl.handle.net/10400.5/3370

Vidal-Suárez, M., & López-Duarte, C. (2013). Language distance and international acquisitions: A transaction cost approach. *International Journal of Cross Cultural Management*, *13*(1), 47–63. https://doi.org/10.1177/1470595812452634

Wentzel, J. P., Diatha, K. S., Seshachal, V., & Yadavalli, S. (2016). *An investigation into factors impacting financial exclusion at the bottom of the pyramid in South Africa*. *33*(2), 203–214. https://doi.org/10.1080/0376835X.2015.1120648

Wentzel, J. P., Diatha, K. S., & Yadavalli, V. S. S. (2016). An investigation into factors impacting financial exclusion at the bottom of the pyramid in South Africa. *Development Southern Africa*, *33*(2), 203–214. https://doi.org/10.1080/0376835X.2015.1120648

Wentzel, J. P., Yadavalli, V. S. S., & Sundar, D. K. (2013). A comparison of financial service adoption propensity across three channels at the Bottom of the Pyramid in South Africa. *Management Dynamics*, *22*(1), 26–38.

Zellner, A., & Lee, T. H. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, *57*(298), 348–368.

Zins, A., & Weill, L. (2016a). The determinants of financial inclusion in Africa. *Review of Development Finance*, *6*(1), 46–57. https://doi.org/10.1016/j.rdf.2016.05.001

Zins, A., & Weill, L. (2016b). The determinants of financial inclusion in Africa. *Review of Development Finance*, *6*(1), 46–57. https://doi.org/10.1016/j.rdf.2016.05.001

Appendices

Appendix A: List of Countries used in SUR Models

Albania
Angola
Argentina
Armenia
Australia
Austria
Bangladesh
Belarus
Belgium
Bhutan
Bolivia, The Plurinational State of
Botswana
Brazil
Bulgaria
Burundi
Cameroon
Chile
China, The People's Republic of
Colombia
Costa Rica
Croatia
Cyprus
Czechia
Denmark
Djibouti
Dominican Republic
Ecuador
El Salvador
Estonia
Eswatini
Finland
France
Gabon
Georgia
Ghana
Guatemala
Honduras
Hungary
Indonesia
Ireland
Italy
Kazakhstan
Kenya
Kyrgyzstan
Latvia
Lesotho
Lithuania
Malawi
Malaysia
Mauritius
Mexico
Moldova, Republic of
Namibia
Netherlands, The
Nicaragua
North Macedonia
Norway
Pakistan
Palestine, State of
Panama
Paraguay
Peru
Philippines, The
Poland
Portugal
Romania
Russian Federation, The
Rwanda
Slovakia
Slovenia
Spain
Sweden
Switzerland
Sri Lanka
Tanzania, The United Republic of
Thailand
Turkey
Uganda
United Arab Emirates
United Kingdom
United States of America
Uruguay
Viet Nam
Zambia