Investigating the Influence of Socioeconomic Characteristics on the Level of Financial Inclusion in Sub-Saharan Africa

MSc Research Project
Fintech

Anthony Falola
Student ID: 20242727

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……. Falola Anthony…………………………………………………………………………
Student ID:	…………20242727………………………………………………………………………
Programme:	……Fintech……………………………… **Year:** ……2021/22.
Module:	……MSC Research Project………………………………
Supervisor:	……Noel Cosgrave……………………………………………………………………. ………
Submission Due Date:	……August 15, 2022……………………………………………………. ………
Project Title:	… Investigating the Influence of Socioeconomic Characteristics on the Level of Financial Inclusion in Sub-Saharan Africa
Word Count:	……8,316………………… **Page Count**……23……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …………………………………………………………………………………………………

Date: …………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Investigating the Influence of Socioeconomic Characteristics on the Level of Financial Inclusion in Sub-Saharan Africa

Anthony Falola
20242727

Abstract

Financial inclusion is critical to reducing income inequality and driving economic development. An inclusive economy means participants can access and use financial products and services. Unfortunately, Sub-Saharan Africa (SSA) is considered one of the least financially inclusive regions of the world. This study uses fixed effects regression to determine which socioeconomic factors, sourced from the World Bank's World Development Indicators (WDI) database, significantly impact inclusion in 22 sub-Saharan African countries. The financial inclusion level of these countries is computed using a two-stage robust principal component analysis (rPCA) based on financial access and usage metrics sourced from the International Monetary Fund's Financial Access Survey database. The study finds that urban population and electricity access have the most significant positive effect on financial inclusion, while an increasing unemployment rate has a negative impact on financial inclusion. Also, changes in the female population ratio negatively influence access to financial services in the region. Decision makers can leverage these insights in policy formulation by better understanding these characteristics. A data unavailability issue limits the number of countries that could be included in the research; however, further work should carry out a cross-country comparison of the impact of socioeconomic characteristics on financial inclusion based on income level.

1 Introduction

1.1 Background of the Study

The drive for increased levels of financial inclusion in any economy is expected to grow the actualisation of economic policy reforms and national transformation. That is why the World Bank categorised financial inclusion as one of the strategic enablers responsible for meeting the United Nation's Sustainable Development Goals of 2030 (Kuada, 2019; World Bank Group, 2021). It has become even more important with the advent of the Covid-19 pandemic that threatened to halt the free flow of business activities worldwide. Moreover, the increased access and usage of financial services for economic participants can serve as a human capital tool toward sustainable economic recovery.

The concept of financial inclusion began in the twenty-first century and is defined as the ability of economic participants, consisting of individuals of legal working age, to use financial services such as basic savings and transactional payments and more advanced insurance and credit services (Sarma, 2008). Consistent usage of financial services bolsters economic activities and is considered a means to reduce the gaps in income inequality. Previous researchers have also identified inclusive finance-led policy reforms as a crucial guide to improved funding allocation, human capital investments and risk management. Inclusive finance allows stakeholders to access financial products and services for various needs (Younas, Qureshi and Al-Faryan, 2022).

The importance of financial inclusion on the economy is well outlined by leading economic organisations and nations such as the World Bank, through various financial inclusion publications and the G20, by committing to implementing its high-level principles for increased digital financial inclusion. More than 60 countries have commissioned specialised financial inclusion strategies since 2010[1], and research in the domain has increased over the years, measuring the results of financial inclusion efforts. However, most of these studies focused on the impact of financial inclusion on various areas of the financial economy, not the influence of the determinants. One such study investigated the impact of an inclusion strategy on economic growth and development with a specific focus on the Middle Eastern economy of Jordan (Alnabulsi and Salameh, 2022).

Sub-Saharan Africa (SSA) remains one of the inclusion-challenged regions of the world, with rising levels of poverty and income inequality (Amponsah, Agbola and Mahmood, 2021). This justifies the focus and effort placed by different decision-makers at various policy formulation levels on improving SSA's financial access and usage. However, it has also been unclear how best to measure the results of the efforts in the region because of the quality of information available to measure inclusion (Abdulmumin *et al.*, 2019). Some researchers have identified ways of measuring the index of financial inclusion in SSA; however, macro and micro determinants' level of influence in driving inclusion in the region has not been extensively studied.

Some vital macro-level determinants influencing financial inclusion on a global scale include GDP per capita, education, regulation, gender, age, and barriers to access (Bekele, 2022). Their level of impact varies depending on the economic state or demography of the economy. Previous literature has measured financial inclusion by deducing secondary-level demand and supply side data sourced by international financial organisations such as the International Monetary Fund[2] and the World Bank[3] to create a multidimensional generated single number using machine learning algorithms. The single number allows for the comparison of financial inclusion across geographies and times.

This study aims at bridging the knowledge gap between the level of financial inclusion and various socioeconomic determinants in sub-Saharan Africa. The research seeks to contribute to the literature by modelling the relationship between socioeconomic determinants and financial inclusion. This will help policymakers develop solutions that can increase financial inclusion in SSA. Furthermore, the result will assist policymakers with comprehensive insights on the importance of these determinants to the drive for increased inclusion in the region.

---

[1] https://www.worldbank.org/en/topic/financialinclusion/overview
[2] https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C
[3] https://www.worldbank.org/en/publication/globalfindex

1.2 Research Question and Objectives

The research question for this study is:

'To what extent do changes in socioeconomic characteristics affect the level of financial inclusion in Sub-Saharan Africa?'

The study objectives are:

- Compute a more up-to-date index of financial inclusion for countries in SSA
- Measure the impact of changing socioeconomic characteristics on the access and usage of financial services
- Examine patterns between changes in socioeconomic determinants and overall financial inclusion in sub-Saharan African countries

The research objectives are achieved methodically using robust principal component analysis to compute the financial inclusion indices and fixed effects regression to examine the socioeconomic determinants and their relationship with the financial inclusion index. An ideal investigation into the impact of socioeconomic determinants on the level of financial inclusion in sub-Saharan Africa should include all the countries that constitute the region. However, the unavailability of sufficient information is a limitation; hence experimental analysis is best carried out using countries with minimum viable data entries while using different imputation techniques to compensate for the gap.

The rest of the study is structured to include a comprehensive literature review in section two, comparing previous research relevant to the domain. Sections three and four detail the research methodology implemented and the design specification. Section five contains the implementation, while section six houses the results and accompanying analysis discussion. Finally, the conclusion and recommendations for future work are in section seven.

2 Related Work

2.1 Evaluating Financial Inclusion

A nation's economic backbone is in its financial services sector, which can be viewed through two lenses. The first is financial sophistication which comprises market depth and liquidity, and the second is financial inclusion. The complexity of a financial market can be gauged through the different market depth and liquidity instruments available for use by individuals and institutions. It is relatively easy to state how a financial market has developed over time through metrics such as foreign capital flows, rates, and average transaction costs. On the other hand, financial inclusion refers to the conversion of non-participants in financial activities to active participants through access to banking or other formal financial services (Lenka, 2021; Ifediora *et al.*, 2022). The accurate measure of financial inclusion has split researchers over the years with no known consensus on the appropriate dimensions to consider or computation method to deploy.

The World Bank outlines the access and usage dimension as two proxies to consider in determining the level of financial inclusion. Financial inclusion means individuals can meet their savings, credit, transactions, and insurance needs through formal financial services.

Gaining access through the availability of such services is the first step to increased inclusion. The World Bank outlines the usage of the available services as the second dimension to consider, especially highlighting countries such as Kenya and India, where more than 80% of the adult population are already account holders (World Bank, 2022). The International Monetary Fund (IMF) publishes the financial access survey (FAS), which provides data to measure access and use of financial services. The IMF also outlines the nine indicators approved by the G20 for measuring financial inclusion, such as the number of ATMs and mobile money transactions for every 100,000 adults[2].

The access and usage dimension of measuring financial inclusion was adopted in a study centred around financial inclusion and language heterogeneity to measure the extent to which future time reference could impact the level of inclusion across different countries. The researchers used ordinary least squares (OLS), probit and instrumental variables techniques on data sourced from the World Bank Global Findex dataset and language-FTR data. The access indicators used in this research include account and debit card ownership data. In contrast, usage indicators had the percentage of respondents with savings in a financial institution and debit card usage volumes (Dar and Sahu, 2022).

One significant difference in the financial inclusion dimensions between the research of Dar and Sahu (2022) and the study on bank stability and inclusion by Wang and Luo (2021) is the composition of the index from indicators measured per $1{,}000KM^2$ as against every 100,000 adults. Basing their study on emerging economies, Wang and Luo (2021) found that banking stability increases with growing financial inclusion. They did this by computing a robust index using principal component analysis (PCA) on normalised indicators representing the usage and access dimensions. The study also validated the data suitability by carrying out the Kaiser-Meyer-Olkin test and Bartlett's test of sphericity. The indicators used were sourced from the IMF FAS database.

The direction of literature when two dimensions are involved in creating an index is typically the principal component analysis approach. Usage and access dimensions can be likened to financial demand and financial supply metrics used in a study that investigated the relationship between bank stability and financial inclusion in the Asian region. The number of ATMs and bank branches represented financial supply. In contrast, credit and debit card usage per 1,000 adults represented financial demand from data collected across different countries in the Asian region. Demand and supply data formed the principal components used to create the index before implementing the generalised method of moments (GMM) that was used to test the relationship between the bank stability variables and the financial inclusion indicator. The GMM technique helped control the high heteroskedasticity and autocorrelation in the dataset (Vo, Nguyen and Thi-Hong Van, 2021).

Danisman and Tarazi (2020) used principal component analysis using data from the Global Findex database for two financial inclusion dimensions. Their research was focused on European countries and considered the high inclusion levels by leveraging digital payment information in contrast to the number of bank branches or ATMs. This is like the study on fintech-based financial inclusion and its relationship with bank-risk taking, highlighting the low-risk culture of Islamic banks compared to traditional banks in the Organisation of Islamic Countries (OIC). The researchers deployed the PCA technique to create a fintech-based financial inclusion index from digital finance usage and access variables (Banna, Kabir Hassan and Rashid, 2021).

Some studies have used more than access and usage dimension. Xu (2020) argued for the inclusion of the second level of measurement that captures the quality of usage of financial services. The researcher also called for expanding the scope of measuring inclusion by considering credit facilities and technologies. This approach is countered in some way by Nuzzo and Piermattei (2020). They discouraged using many metrics from large data sources as they can be ineffective in localising financial inclusion. The study argued for the exclusion of credit indicators as financial inclusion should only be determined by access and savings.

One of such additional dimensions considered for the computation of a financial inclusion index is availability. Besong, Okanda and Ndip (2022) split the access dimension by using the number of deposit and loan account holders as indicators for access, while the number of bank branches and ATMs per 100,000 adults described availability. The research investigated the relationship between regulations and inclusion in the financial services sector of the Central African Economic and Monetary Community region. Fixed effect panel regression proved the positive relationship between the financial inclusion index, the dependent variable, and endogenous independent variables. A quality indicator formed part of the metrics used in a study that sought to establish a relationship between institutional business environment and inclusion. The researchers included domestic credit to the private sector as a part of four proxies that formed the variables in a panel vector autoregressive model (PVAR) on sample data of 43 countries. PVAR was preferred for its ability to show the variance composition through each variable's contribution (Charfeddine and Zaouali, 2022).

Some researchers have combined the indicators to form a single index without prioritising dimensionality, using supply-side data on insurance, savings, and penetration. The research found a negative relationship between non-performing loans, financial inclusion and economic growth using the Driscoll-Kraay standard errors with a fixed effect on large panel data from 21 countries from the Organisation for Economic Corporation and Development (OECD) (Zhang *et al.*, 2022). These findings set the theme for adopting the two central dimensions of financial inclusion, namely, access and usage, for this study, considering their consistent presence in the reviewed literature.

2.2 Cross-Country Research on Financial Inclusion

The expectation of increased inclusion varies from developed to developing and underdeveloped economies. Highly sophisticated economies tend to research the effect of changes in inclusion on key economic determinants to drive productivity and innovation. The drive for developing economies is to make changes that can prevent financial crises (Lee, Wang and Ho, 2022). Fareed et al. (2022) posited that inclusion in developed countries is more focused on financial literacy than fundamental participation, whereas developing economies measure financial inclusion levels to improve the insufficient or flawed financial systems. Regardless of the research motive, financial inclusion is as important in developed nations as in developing countries. Research showed income inequality driven by low inclusion levels in some parts of the United States, one of the most developed nations in the world. The region had 79% adult account ownership compared to Japan, a similarly sophisticated economy with 98% account ownership (Kumar, Thrikawala and Acharya, 2021).

To understand the properties of inclusion in developing areas, researchers such as Marcelin et al. (2021) developed a dataset of 44 developing countries because of their weak institutions, inflation, and competition to establish the relationship between financial inclusion and economic performance using GDP per capita. The range and selection of countries are similar to that of Wang and Luo (2021), who used panel data of 36 emerging countries across Latin America, Europe and Asia. Despite being a hotbed of developing economies, their research had no African representation. One of the researchers that considered Africa investigated the role of inclusion in the relative volatility of gross capital formation. The data was sourced from the Middle East and North Africa because of the region's financial insufficiencies, such as limited securities for credit (Rojas Cama and Emara, 2022).

The World Bank reports that only 24% of SSA's adult population had a bank account as of 2011 (World Bank Group, 2021). Despite this, few financial inclusion studies have been based in the region. One is a study to investigate patterns between income inequality, institutional quality and financial inclusion using a finite mixture inequality model. The research questioned the expected impact of increasing inclusion on inequality as the SSA dataset displayed inequality regimes not previously seen in literature focused on other regions (Sawadogo and Semedo, 2021). The research represented further work on initial financial inclusion studies in the area. Due to the peculiarities of missing data in SSA, Abdulmumin et al. (2019) could only use 22 SSA economies to determine the level of financial inclusion in the region. Their research highlighted Seychelles as the most financially inclusive country in the region based on the index of financial inclusion created using principal component analysis.

Amponsah, Agbola and Mahmood (2021) combined data from the IMF's financial access survey (FAS) with other global data sources such as the World Bank World Development Indicators and World Governance Indicators for 44 countries in SSA. The research was centred around informality, growth, and financial inclusion. The study did not assign dimensions to the financial inclusion indicators in the PCA model; instead, using an equal weight on all the data variables. The FAS database was equally helpful in an unsupervised machine learning technique deployed to measure financial inclusion in relation to the banking market structures. The researchers used two-stage PCA by assigning the indicators to the availability, access, and usage dimensions. The two-stage PCA was effective in handling the over-parameterisation of data. One limitation of this study is that only 17 African countries were used (Kebede, Naranpanawa and Selvanathan, 2021).

The two-stage PCA process proved useful for Nguyen (2021), who attempted to create a composite index for the measurement of financial inclusion using data for 40 developing economies across Africa, Oceania, and Asia. The first stage model sought the indicators into the usage, access, and availability dimensions before the computation of the financial inclusion index. The work of Khmous and Besim (2020), based on data from 14 MENA countries, is one of the few studies in the financial inclusion domain that used other techniques for their research. A probit estimation technique that used binary indicators relating to usage and barriers to inclusion found that Islamic banking practices were one of the critical factors responsible for the lower average inclusion level of middle-income MENA countries.

2.3 Identifying Key Socioeconomic Characteristics

When the financial inclusion domain research has used primary sourced data, researchers can outline determinants to be tested against a financial inclusion index. For example, a rural-based study used binary logistic regression on a 780-household sample size in India to determine the relationship between financial inclusion and some socioeconomic factors. The researchers used financial literacy, income status and the primary income source of the household as variables to prove the vulnerability of certain socioeconomic groups in financial inclusion (Kandari, Bahuguna and Salgotra, 2021). Mouna and Jarboui (2021) had similar results despite using secondary data. In addition, a probit estimation model helped prove the significant influence of variables such as age, level of education and gender on the level of financial inclusion in MENA countries.

In examining economic growth and inclusion from a non-linear standpoint, a paper included variables such as human capital index, annual growth rate, population growth rate, inflation rate, unemployment rate and trade openness. This was to show the low impact an increasing financial inclusion standard had on the economy in areas of high inclusivity (Abdul Karim *et al.*, 2021). The variables of choice are very different from those used by a paper aimed at discovering the influence of individualism on household financial inclusion. The researchers gathered country-specific data from the World Bank's Global Findex report of 2014. The variables considered include age, gender, income level of the participants, real GDP per capita and level of education. The findings reveal a stronger positive relationship between individualism and financial inclusion, especially within the lower socioeconomic class of the sample (Lu, Niu and Zhou, 2021).

The choice of determinants and the extent of their relationship has been influenced by the focus area's geographical location and economic status. Danisman and Tarazi (2020) identified the need to characterise these determinants as a means to direct policy attacks on financially exclusive groups. Their research, focusing on the European Union (EU), identified gender, rural to urban distribution, education level, employment status and age as a basis for financial stability in the EU. The young under-educated adult group was identified as the weak link to financial inclusion in the research. The determinants identified are similar to the group of socioeconomic indicators used in a polychoric principal component analysis and OLS on data from South African consumers. The research sought to measure a quality financial inclusion metric in relation to gender, education, financial literacy, employment status and income level. Demographic and access indicators such as bank distance and geographical location also formed part of the independent variables of the research (Chipunza and Fanta, 2021).

3 Research Methodology

The research techniques employed for this study are detailed in this section in line with the Knowledge Discovery in Database (KDD) approach. The section contains data selection details, research procedure and techniques.

3.1 Data Selection

The IMF Financial Access Survey, a supply-side secondary data source on financial inclusion indicators, was used for the analysis. The IMF FAS contains financial inclusion and other economic indicators collected annually for 189 countries. Data for 22 Sub-Saharan African countries over ten years from 2011-2020 was used for the empirical analysis. There are 49 countries in SSA[4]; however, due to the unavailability of financial inclusion data for many countries, the research had to be limited to only countries with available data. In addition to the

[4] https://openknowledge.worldbank.org/pages/focus-sub-saharan-africa

financial inclusion indicators, socioeconomic metrics were sourced from the World Bank's World Development Indicators (WDI) database[5].

The IMF FAS dataset was favoured over its combination with the Global Findex dataset justified by literature such as Nuzzo and Piermattei (2020) because the Global Findex dataset is triannual, limiting the number of observations available for analysis. Therefore, the final dataset used for the study was composed of two main segments of variables in addition to the panel data structure for Sub-Saharan African countries, namely the Index of Financial Inclusion (IFI) Indicators and the socioeconomic indicators.

3.1.1 Index of Financial Inclusion

Central banks and financial regulators collect the IMF FAS dataset for each country with a comparative analysis of access to financial services and the level of usage of the services rendered[6]. For this research, the indicators proposed by the G20 and used by Vo, Nguyen and Thi-Hong Van (2021) and Dar and Sahu (2022) for measuring financial inclusion are used.

The access dimension represented as penetration and availability in some literature refers to the reach of financial services and their enablers to the economic participants. The access indicators used in this study include the number of commercial bank branches for every 100,000 adults, the number of Automated Teller Machines (ATM) for every 100,000 adults and the number of registered mobile money agent outlets for every 100,000 adults.

Usage indicators such as the number of mobile money transactions per 1,000 adults, number of deposit account with commercial banks per 1,000 adults, and number of loan accounts with commercial banks per 1,000 adults, all part of the proposed G20 financial inclusion indicators were used. The usage index represents the extent to which financial services are used over time[7]. However, the usage indicators concerning small and medium-sized enterprises (SMEs) were left out of the study due to many missing data points.

3.1.2 Socioeconomic Indicators

The data on socioeconomic variables have been retrieved from the World Bank's World Development Indicators (WDI) database, a collection of global development data from officially accredited sources. The choice of the variables is based on the key determinants of financial inclusion identified in past literature. For this study, the percentage of the urban population, access to electricity, share of the people with internet access, the rate of unemployment, the female population, and the number between the working age of 15 to 64 in the country. In addition, GDP per capita was included as a control variable for its recognition as an indicator of economic growth (Kim, Yu and Hassan, 2018; Zhang *et al.*, 2022).

---

[5] https://databank.worldbank.org/source/world-development-indicators

[6] https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C

[7] https://www.worldbank.org/en/topic/financialinclusion/brief/how-to-measure-financial-inclusion

Table 1: Definition of selected variables

Variable name	Definition
UrbanPopulation	The total number of people living in urban areas as a percentage of the total population
ElectricityAccess	The percentage of the population with access to electricity
InternetUsage	Number of individuals using the internet as a percentage of the total population
Unemployment	The estimated rate of the total available labour force that is seeking employment
FemalePopulation	The number of females as a percentage of the total population
WorkingAge	The number of people aged between 15 and 64 compared to the total population
GDPpercapita	The gross domestic product of each country per individual at the current US dollar conversion rate
UsageIndex	Index prepared using PCA, including the number of mobile money transactions per 1,000 adults, number of deposit accounts with commercial banks per 1,000 adults, number of loan accounts with commercial banks per 1,000 adults
AccessIndex	Index prepared using PCA, which includes the number of commercial bank branches for every 100,000 adults, the number of Automated Teller Machines (ATM) for every 100,000 adults and the number of registered mobile money agent outlets for every 100,000 adults
IFI	The index includes the proxies for the access and usage dimensions of financial inclusion

3.2 Pre-processing

This stage involves exploratory data analysis on the selected datasets to tell what the data looks like and what can be done. Pre-processing was carried out using R studio after importing the IMF FAS and WDI datasets from online sources. Asides from descriptive statistics, dealing with missing data points with different imputation techniques were performed on both datasets. Pre-processing also involved the removal of unwanted observations or variables, dealing with outliers and data standardisation.

3.3 Data Transformation

Data transformation involves the reordering of columns, merging of datasets and the application of the variables to fit the model specification. Data transformation was carried out separately for both datasets before creating a project dataset, a product of the results from the financial inclusion analysis and the socioeconomic variables extracted from the WDI database. The variables were also renamed to make data mining less prone to error and results easier to interpret.

3.4 Data Mining

Data mining seeks to provide information from data. This project deployed robust principal component analysis to analyse financial inclusion data and derive a financial inclusion index and fixed effects regression to establish the relationship between the derived financial inclusion index and socioeconomic variables.

3.4.1 Robust Principal Component Analysis

The principal component analysis is common in the financial inclusion domain as a preferred multivariate method of computing financial inclusion index using principal components. The principal components result from the orthogonal transformation of correlated variables to linearly uncorrelated variables while retaining as much information as possible. Since the main task PCA seeks to achieve is the reduction of dimensionality in the data with the minimum mean square error, it creates a new coordinate system by finding an orthonormal base for the variables (Abdulmumin *et al.*, 2019; Mahmoudi *et al*., 2021; Li *et al.*, 2022).

PCA was developed by Karl Pearson in 1901 and has subsequently been developed by other researchers such as Harold Hotelling and Karhunen Loeve (Younes *et al.*, 2021). One of such variations to PCA is the robust Principal Component Analysis which can work with missing data points and significant outliers. rPCA maximises subspace recovery and uses matrix completion to accommodate data points that can be considered helpful for the research (Vaswani, Chi and Bouwmans, 2018). rPCA is favoured for the computation of financial inclusion in Africa due to the likelihood of outliers, missing data entries and the multicollinearity of the variables.

3.4.2 Fixed Effects Regression

The fixed effects model is a regression model that analyses the relationship between variables in a panel dataset. Fixed effects models are also referred to as the within estimator, investigating the impact of the independent variables on the dependent variable for each defined entity in the panel data over a time series. Fixed effects models remove the effect of time-invariant properties by de-meaning the variables and checking the relationship within the clusters relative to the mean (Bell *et al.*, 2019). Theoretically, fixed effects models perform better than random effects models in cases of unobservable entity effects across unobservable time effects. The preference for the fixed effects model is to accommodate the unobservable heterogeneity considering various countries display different characteristics and can be influenced by multiple factors (Songwathana, 2018; Jarboui, 2021).

4 Design Specification

The model specification implemented in this research is defined as

$$IFI_{it} = f(X1_{it}, X2_{it}, X3_{it}, \dots, Xn_{it}) \qquad \textbf{(1)}$$

Where $X1_{it}$…, Xn_{it} in equation 1 represents the independent variables for each country i and at year, t. The IFI is derived by robust principal component analysis and serves as the dependent variable in the panel regression equation defined by Baltagi (2001) as:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \qquad \textbf{(2)}$$

Where X_{it} represents the independent variables, β is the coefficient for the independent variable α and ε is the intercept and the error term respectively.

4.1 Robust Principal Component Analysis

The purpose of the financial inclusion index is to capture the various dimensions of financial inclusions in comparative values for the different entities (i) across the time (t). Therefore, a two-stage rPCA technique carries out indexing of the indicators into the dimensions in the first stage before collapsing the dimensions into the final IFI in the second stage. This method is similar to the approach of Cámara and Tuesta (2014), where a two-stage PCA was implemented to accommodate the financial inclusion indices due to the weight bias of PCA towards highly inter-correlated indicators.

The conclusion of the first stage of rPCA provides access to the financial services index (*AccessIndex)* and the usage of financial services index (*UsageIndex*). Both indices serve as explanatory variables for the development of the IFI. The IFI is linearly derived in equation 3 by

$$IFI_{it} = \omega_1 AccessIndex_{it} + \omega_2 UsageIndex_{it} + \varepsilon_{it} \tag{3}$$

The error term ε is expected to show slight variance considering that the variation in the variables included in the access and usage dimension will largely explain the variation in financial inclusion (IFI).

4.1.1 First Stage Robust Principal Component Analysis

The first stage of rPCA develops the access and usage dimension index from the selected indicators with the equations as follows:

$$AccessIndex_{i,t} = \beta_1 ATM.pop_{i,t} + \beta_2 Branch.pop_{i,t} + \beta_3 MMOutlet_{i,t} + u_{i,t} \tag{4}$$

$$AccessIndex_{i,t} = \frac{\sum_{j,k=1}^{t} \lambda_j P_{k,i,t}^{a}}{\sum_{j,k=1}^{t} \lambda_j^{a}} \tag{5}$$

Where *ATM.pop* is the number of ATMs for every 100,000 adults, *Branch.pop* is the number of commercial bank branches for every 100,000 adults, and the *MMOutlet* is the number of mobile money outlets per 100,000 adults. The access dimension is derived as a linear function of these three indicators. Equation 5 represents the weighted average estimation where λ_j is the jth eigenvalue for the dimension. The j is the number of indicators which form the principal components, while Pk denotes the kth principal component. $Pk_{i,t}$ is provided by the rPCA scores and includes all the components for the dimension as only three indicators were used; hence, the need for dimensionality reduction is minimal. A similar approach is carried out for the usage dimension using *MMT*, which is the number of mobile money transactions per 1,000 adults, *Dep.Acct* denoting the number of deposit accounts with commercial banks per 1,000 adults and *Loan.Acct*, which is the number of loan accounts with commercial banks per 1,000 adults, as its indicators.

$$UsageIndex_{i,t} = \theta_1 Dep.Acct_{i,t} + \theta_2 Loan.Acct_{i,t} + \theta_3 MMT_{i,t} + v_{i,t} \tag{6}$$

$$\text{UsageIndex}_{i,t} = \frac{\sum_{j,k=1}^{t} \lambda_j P_{k,i,t}^{u}}{\sum_{j,k=1}^{t} \lambda_j^{u}} \tag{7}$$

4.1.2 Second Stage Robust Principal Component Analysis

A similar indexing approach followed in the first stage is repeated in the second stage with the AccessIndex, and UsageIndex used as the indicators instead, as in equation 3. Robust PCA calculates the corresponding weight of each dimension as shown in equation 8. λ_j represents the eigenvalue at j = 1 and 2. ψ is the eigenvectors of the correlative matrices of the various dimensions.

$$\omega_k = \frac{\sum_{j=1}^{2} \lambda_j \psi_{jk}}{\sum_{j=1}^{2} \lambda_j} \quad where\ k = 1\ and\ 2 \tag{8}$$

The eventual *IFI* is derived as a value between 0 and 1 for each entity at period t and is represented by equation 9 below.

$$\mathrm{IFI}_{i,t} = \frac{\sum_{j=1}^{2} \lambda_j (\psi_{j1} AccessIndex_{i,t} + \psi_{j2} UsageIndex_{i,t})}{\sum_{j=1}^{2} \lambda_j} \tag{9}$$

4.2 Fixed Effects Regression

The key difference between the fixed effect model and the panel data regression model in equation 2 is the unknown intercept αadapted to each entity in a fixed effect model as against being constant. The FE model for this research is as follows:

$$\begin{aligned} IFI_{it} = {} & \alpha_i + \beta_1 UrbanPopulation_{it} + \beta_2 ElectricityAccess_{it} + \beta_3 InternetUsage_{it} + \\ & \beta_4 Unemployment_{it} + \beta_5 FemalePopulation_{it} + \beta_6 WorkingAge_{it} + \\ & \beta_7 InGDPpercapita_{it} + \varepsilon_{it} \end{aligned} \tag{10}$$

IFI is regressed on an intercept, αi, which is specific to the entity(country) and the other independent variables.

5 Implementation

5.1 Data Selection

This research utilised secondary data sources for the financial inclusion data and socioeconomic variables from the IMF's FAS database and the World Bank's WDI database. The IMF's FAS dataset was filtered for Sub-Saharan African countries and downloaded in csv format before import into R studio. At import, the FAS dataset contained 720 observations and 194 variables ranging from 2004 to 2021. The WDI dataset was downloaded more specifically on the focus countries based on the completeness of data available for the entities in the FAS dataset. At the point of import into R studio, the WDI dataset had 247 observations of 12 variables ranging from 2010 to 2020.

5.2 Pre-processing and Transformation

To effectively answer the research question, the study used a balanced panel dataset that accounted for uniform observations in all entities across a particular time frame. In addition, some pre-processing data procedures were carried out on each dataset before merging the data.

5.2.1 FAS Dataset

The first data wrangling step in the FAS dataset was the selection of the appropriate indicators for the research. Upon missing data analysis and manual inspection, the countries with a notoriety of missingness were removed from the study. Also, the data suggested missing values for mobile money indicators across various economies. At the same time, the IMF's FAS metadata indicates that mobile money had only been introduced in countries like Chad and Mauritius in 2013, Seychelles in 2014, the Gambia and Equatorial Guinea in 2016, and mobile money services are non-existent in South Sudan. Therefore, these missing entries were replaced with 0. Asides from the Economy variable, which depicts the country, and the Year variable, all other variables were converted to numeric. After carrying out these cleaning procedures, the percentage of missing entries in the dataset was 8.4%, necessitating some imputation techniques.

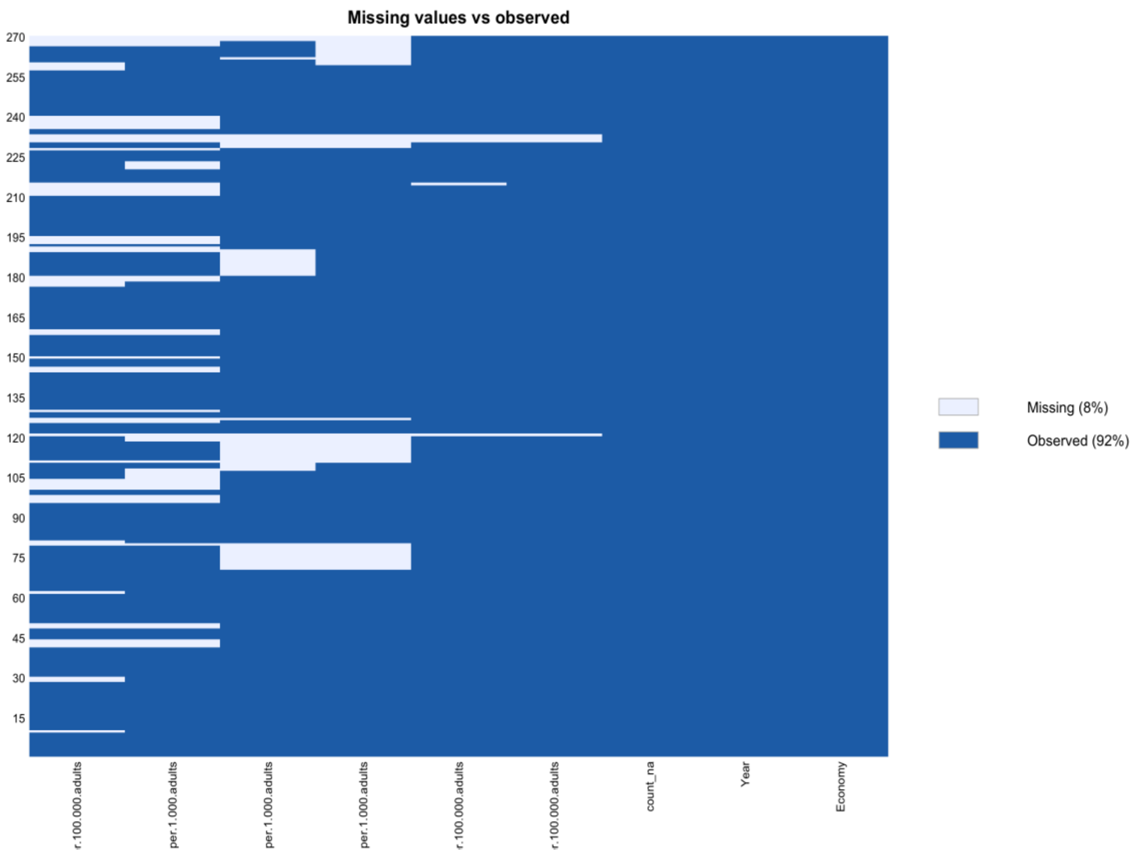

Figure 1: Boxplot of missing data for FAS dataset

The next phase of the FAS dataset clean-up involved using the last observation carried forward (LOCF) and mean substitution. These techniques were used in instances where data was available in the prior and or the following year for the same variable in the same country (de Goeij *et al.*, 2013). All the columns were reordered, and the variables were renamed appropriately to reduce the odds of error. To completely address the outstanding missing data points, Probabilistic Principal Component Analysis (PPCA), a technique that assumes the independence of data points in latent variables where data is missing at random, was used (Ghojogh *et al.*, 2022). PPCA had outperformed Multiple Imputation using Chained Equation (MICE) in mitigating the impact of missing data. It uses Expectation-Maximation (EM) algorithm to estimate values by recovering original data from a dimension-reduced form (Hegde *et al.*, 2019).

5.2.2 WDI Dataset

The 247 observations in the WDI dataset were pruned down to match the number of observations in the FAS dataset by removing the unwanted rows such as data for 2010 for the observed countries and other empty rows. The country and year code columns were removed while the columns were renamed appropriately. The WDI dataset had two dots in missing data entries which were substituted with NA for a correct missing data analysis to be carried out. The unemployment rate data for Seychelles was missing entirely and replaced by the national estimates for each year from the same data source.

The primary school age variable, representing the number of children enrolled in primary education each year, was completely removed from the dataset as it had 28% missing values. The remaining missing values were treated after merging the WDI dataset with the results of the rPCA on the FAS dataset.

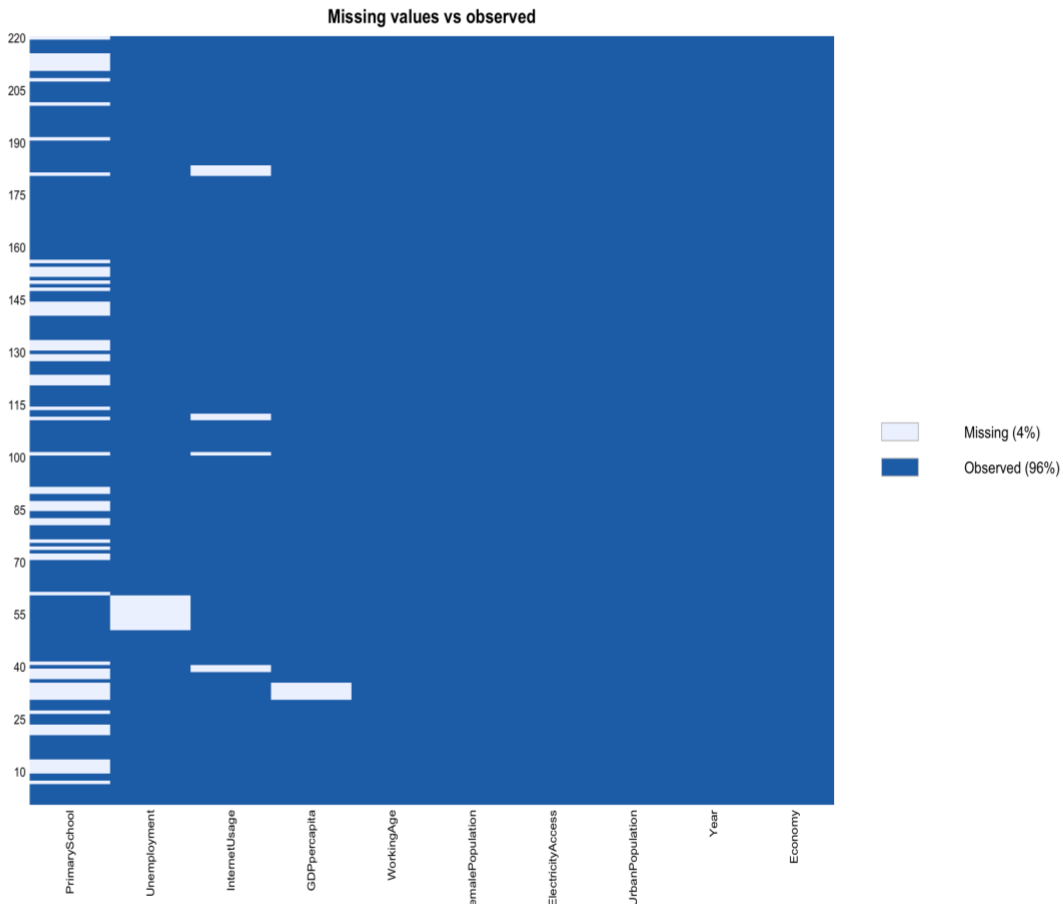

Figure 2: Boxplot of missing data for the WDI dataset

5.2.3 Index of Financial Inclusion (IFI)

The cleaned FAS dataset was subject to Bartlett's test of sphericity and the Kaiser-Meyer-Olkin's (KMO) factor adequacy to test its suitability for PCA. The p-value is lower than 0.05, which shows that the relationship between the variables diverges from the identity matrix. It is not totally uncorrelated, preventing PCA from obtaining a representation of the dimensions. The KMO test produced an overall measure of sampling adequacy of 0.83, suggesting a significant measure of variance can be explained by the common variance, making it suitable for dimension reduction.

The first two dimensions were derived using the rPCA from the 'PcaCov' function in line with equations 4 and 6. Equation 5 and 7 represents the final computation of the dimension indices. It is derived by scaling the values obtained from the principal components of the countries for each year, substituting these scores and the eigenvalues in the equation. The resulting values are the "UsageIndex" and "AccessIndex", which form the two variables for the second stage of rPCA. A similar process is carried out in line with equation 9, which produces the IFI for each Economy and Year.

Table 2: Mean of indices for Sub-Saharan Africa economies

Country	AccessIndex	Rank	UsageIndex	Rank	IFI	Rank
South Africa	0.758	1	0.699	1	0.915	1
Mauritius	0.399	3	0.424	2	0.368	2
Namibia	0.408	2	0.414	3	0.361	3
Botswana	0.375	5	0.402	4	0.327	4
Kenya	0.348	9	0.395	5	0.302	5
Eswatini	0.379	4	0.378	6	0.301	6
Ghana	0.356	6	0.377	8	0.286	7
Zimbabwe	0.345	11	0.378	7	0.280	8
Nigeria	0.348	10	0.374	9	0.277	9
Lesotho	0.350	7	0.367	11	0.270	10
Uganda	0.341	13	0.369	10	0.267	11
Rwanda	0.349	8	0.364	12	0.266	12
Gambia	0.342	12	0.362	14	0.259	13
Cameroon	0.338	14	0.362	13	0.257	14
Malawi	0.336	16	0.359	17	0.253	15
Guinea	0.336	17	0.359	16	0.252	16
Zambia	0.337	15	0.358	19	0.251	17
Liberia	0.334	18	0.359	18	0.250	18
Madagascar	0.331	19	0.360	15	0.250	19
Chad	0.328	20	0.355	20	0.243	20
Seychelles	0.326	21	0.354	21	0.240	21
South Sudan	0.324	22	0.353	22	0.238	22

5.2.4 Merged Dataset

The new variables derived by the rPCA were merged with the WDI dataset to form a dataset of 220 observations and 12 variables. Mean substitution, LOCF and linear regression imputation techniques were used to eliminate all missing data entries. In the case of the GDPpercapita metrics for South Sudan, the missing entries were replaced by the values from the United Nations Database[8]. All variables in percentages were converted to numbers, while the structure of all the variables except the economy and year was changed to numeric. Finally, boxplots of the indices scores were plotted to check for outliers in the merged dataset.

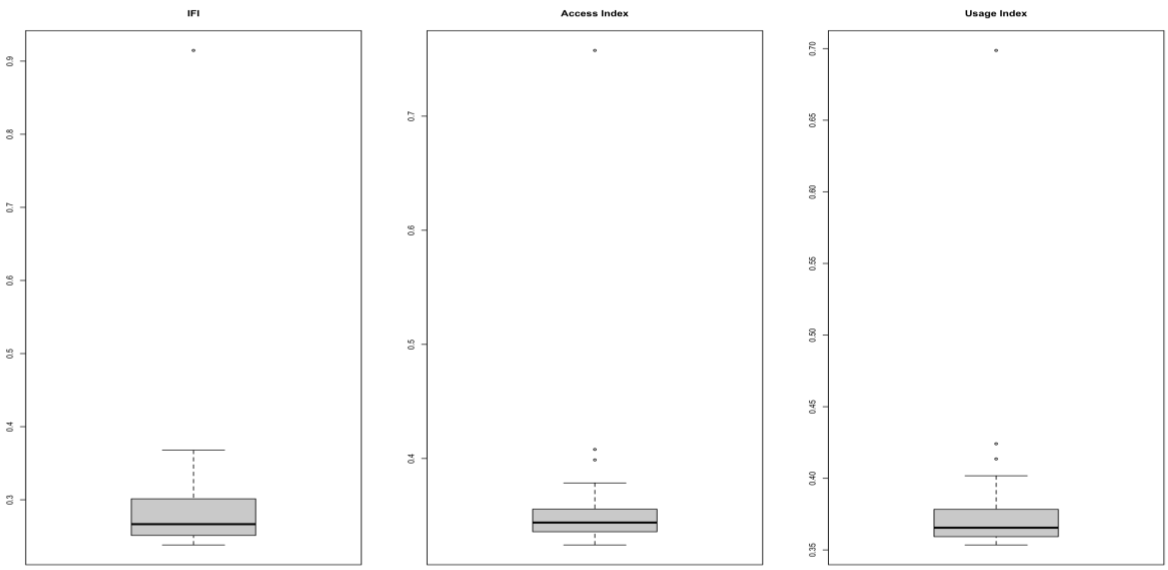

Figure 3: Boxplot of the index scores

5.2.5 Correlation

The correlation among the variables is checked with the results presented in Table 3. IFI is positively correlated with all the variables. At the same time, the only test of negative correlation was the relationship between *FemalePopulation* and a trio of other variables, namely *UrbanPopulation*, *ElectricityAccess* and *InternetUsage*.

Table 3: Correlation Matrix

	UrbanPopulation	ElectricityAccess	InternetUsage	Unemployment	FemalePopulation	GDPpercapita	WorkingAge	UsageIndex	AccessIndex	IFI
UrbanPopulation	1.000									
ElectricityAccess	0.628	1.000								
InternetUsage	0.569	0.846	1.000							
Unemployment	0.255	0.277	0.380	1.000						
FemalePopulation	-0.230	-0.177	-0.140	0.378	1.000					
GDPpercapita	0.467	0.763	0.748	0.256	-0.218	1.000				
WorkingAge	0.414	0.775	0.805	0.382	-0.163	0.842	1.000			
UsageIndex	0.397	0.432	0.470	0.519	0.107	0.334	0.459	1.000		
AccessIndex	0.393	0.418	0.481	0.561	0.119	0.316	0.443	0.977	1.000	
IFI	0.398	0.429	0.477	0.539	0.112	0.329	0.455	0.996	0.992	1.000

[8] https://data.un.org/default.aspx

5.3 Data Mining and Validation

5.3.1 Fixed Effects Regression

The data modelling was implemented using the R studio platform. To successfully carry out fixed effects regression, the PLM package had to be installed and the data classified as panel data outlining the index columns. Then, three estimation models were run on the data with the same independent variables but using the IFI, AccessIndex and UsageIndex, as the dependent variables. The model carries out the regression plot assuming the slope is constant, but the different countries have varying intercepts, as shown in equation 10. Finally, the FE model is carried out on the AccessIndex and UsageIndex to compare the results with the model on the financial inclusion index.

5.3.2 Hausman Test

Hausman is a test necessary for detecting potential endogeneity among regressors. In this study, Hausman compares the consistency of the fixed effect model in contrast to a random effects model. An alternative hypothesis using the phtest() function shows the preference for the fixed effects model because the random effect estimator is inconsistent (Aït-Sahalia and Xiu, 2019).

5.3.3 VIF Test for Multicollinearity

The variance inflation factor (VIF) is used to analyse the dataset for multicollinearity. The purpose of carrying out the multicollinearity test is to check variable redundancy because of the correlation between two or more independent variables. Therefore, the presence of multicollinearity could influence the regression coefficients. There is no formal agreement on the acceptable VIF value, but there is an understanding in previous literature that VIF below 10 remains acceptable (Yoo *et al.*, 2014).

5.3.4 Test for Heteroscedasticity

A Breusch-Pagan test is carried out on the models using the bptest() function from the lmtest library. A null hypothesis confirms homoscedasticity; however, should the null hypothesis be rejected, heteroscedasticity is best controlled using the robust covariance matrix on the model's coefficients.

6 Evaluation of Results, Discussion and Limitation

The results of the fixed effect models for the access and usage dimension and the index of financial inclusion (IFI) are shown below

Table 4: Fixed effects model results

		Inclusion		
		Access	**Usage**	**IFI**
UrbanPopulation	*co-efficient*	0.297**	0.450**	0.716**
	p-value	0.050	0.025	0.013
	standard error	(0.150)	(0.199)	(0.287)
ElectricityAccess	*co-efficient*	0.045	0.173***	0.285***
	p-value	0.118	0.000	0.000
	standard error	(0.029)	(0.038)	(0.055)
InternetUsage	*co-efficient*	0.074***	(0.087)***	(0.038)
	p-value	0.001	0.003	0.362
	standard error	(0.022)	(0.029)	(0.041)
Unemployment	*co-efficient*	(0.293)**	(0.255)	(0.534)**
	p-value	0.015	0.107	0.020
	standard error	(0.119)	(0.158)	(0.228)
FemalePopulation	*co-efficient*	(2.959)***	0.620	(1.129)
	p-value	0.000	0.546	0.451
	standard error	(0.779)	(1.034)	(1.489)
log(GDPpercapita)	*co-efficient*	(0.000)	0.007	0.008
	p-value	0.967	0.398	0.495
	standard error	(0.006)	(0.008)	(0.011)
WorkingAge	*co-efficient*	0.105	0.158	0.523
	p-value	0.696	0.655	0.312
	standard error	(0.267)	(0.355)	(0.511)
R-Square		0.538	0.286	0.517
F-statistics		31.825***	10.926***	29.2623***
p-value		2.22E-16	1.41E-11	2.22E-16
Note:	**p<0.1; **p<0.05; ***p<0.01*			

6.1 IFI results

In the model results presented in Table 4, socioeconomic indicators such as UrbanPopulation, ElectricAccess and WorkingAge had a positive effect on IFI while InternetUsage, Unemployment and FemalePopulation had a negative effect. UrbanPopulation, which is the total number of people living in the Urban areas and ElectricityAccess, representing the number of people with access to electricity in the economy, had the most significant effect on IFI at a 1% significance level. This means that for every 1% increase in the number of people living in the urban cities compared to the total population, the financial inclusion level of the economy increased by 0.72%. On the other hand, an increase in the unemployment rate decreased the financial inclusion level. The model results for this negative effect of 0.534% were also statistically significant. Overall, the F-statistics reports a probability below 5%, confirming the model's validity.

6.2 Access and Usage Index

The fixed effects model also had statistically significant F-statistics for both models used to test the relationship between changing socioeconomic characteristics and the dimensions of financial inclusion in the study. UrbanPopulation, ElectricityAccess and WorkingAge had positive effects on both dimensions. While InternetUsage had a positive impact on access to financial services, there was a statistically significant negative effect on the usage of financial products, with a 1% increase in the number of people who use the internet as a percentage of the entire population reduced the financial inclusion levels by 0.087%. A noteworthy observation is a significant and negative effect an increase in the rate of females in the population had on access to financial services while increasing the usage index at a significant level. A 1% increase in FemalePopulation reduces access to financial services by 2.96%.

6.3 Model Validation and Assumptions

The result of the VIF analysis for multicollinearity of the study variables is shown in Table 5 below.

Table 5: VIF analysis

	IFI	
	VIF	**1/VIF**
UrbanPopulation	1.922	0.520
ElectricityAccess	5.899	0.170
InternetUsage	4.678	0.214
Unemployment	1.834	0.545
FemalePopulation	1.458	0.686
log(GDPpercapita)	4.407	0.227
WorkingAge	4.155	0.241

None of the variables had a variance inflation factor that is considered too high as many of the variables were moderately correlated. However, the ElectricityAccess variable had a relatively high correlation of 5.899, indicating that the coefficient variance could be much higher than expected.

The results of the Hausman test in Table 6 confirm the preference of the fixed effects model over the random effects model due to data endogeneity. The null hypothesis, which favours the exogeneity of individual random effects, is rejected, considering the p-value is less than 0.05.

Table 6: Hausman test

	AccessIndex	**UsageIndex**	**IFI**
Chi-Sq Statistics	21.451	6.897	26.734
Prob.	0.00316	0.43970	0.00037

Table 7 shows a test for heteroscedasticity at a 1% level. For all three models, the homoscedasticity null hypothesis is not rejected using the Breusch-Pagan test. A robust covariance matrix is performed on the fixed effects model to control for heteroscedasticity, and the adjusted results are shown in Table 8.

Table 7: Test for heteroscedasticity

Breusch-Pagan	Dependent Variable	p-value
Model 1	AccessIndex	6.08E-15
Model 2	UsageIndex	2.37E-13
Model 3	IFI	1.02E-14

The coefficients from the test results after the robust covariance matrix remain similar to the initial results of the model. The only change is the increased significance level of the positive effect of ElectricityAccess to the AccessIndex.

Table 8: Result after controlling autocorrelation and heteroscedasticity

		Inclusion		
		Access	**Usage**	**IFI**
UrbanPopulation	*co-efficient*	0.297**	0.451**	0.726**
	p-value	0.021	0.050	0.039
	standard error	(0.150)	(0.228)	(0.349)
ElectricityAccess	*co-efficient*	0.045*	0.173***	0.288***
	p-value	0.078	0.000	0.000
	standard error	(0.025)	(0.035)	(0.055)
InternetUsage	*co-efficient*	0.074***	(0.087)***	(0.038)
	p-value	0.003	0.007	0.424
	standard error	(0.024)	(0.031)	(0.048)
Unemployment	*co-efficient*	(0.293)**	(0.257)	(0.542)**
	p-value	0.020	0.120	0.016
	standard error	(0.124)	(0.164)	(0.222)
FemalePopulation	*co-efficient*	(2.959)***	0.627	(1.137)
	p-value	0.000	0.457	0.366
	standard error	(0.556)	(1.034)	(1.253)
log(GDPpercapita)	*co-efficient*	(0.000)	0.007	0.008
	p-value	0.943	0.128	0.242
	standard error	(0.003)	(0.004)	(0.007)
WorkingAge	*co-efficient*	0.105	0.160	0.525
	p-value	0.524	0.544	0.202
	standard error	(0.164)	(0.262)	(0.409)
Note:	**p<0.1; **p<0.05; ***p<0.01*			

Lastly, considering the indices for South Africa were outliers compared to the rest of the countries considered in the study, the fixed effects model was tested on the dataset while excluding South Africa. The results, as shown in Table 9, only materially differ from the initial model's outcome by the statistical insignificance of the negative effect of unemployment on financial inclusion and its dimensions.

Table 9: Fixed effects model result after removing outliers

		Inclusion		
		Access	**Usage**	**IFI**
UrbanPopulation	*co-efficient*	0.221*	0.461**	0.696**
	p-value	0.064	0.011	0.017
	standard error	(0.118)	(0.179)	(0.287)
ElectricityAccess	*co-efficient*	0.031	0.151***	0.261***
	p-value	0.178	0.000	0.000
	standard error	(0.023)	(0.035)	(0.055)
InternetUsage	*co-efficient*	0.094***	(0.063)***	(0.006)
	p-value	0.000	0.018	0.885
	standard error	(0.017)	(0.031)	(0.042)
Unemployment	*co-efficient*	(0.102)	(0.186)	(0.385)
	p-value	0.298	0.210	0.107
	standard error	(0.098)	(0.147)	0.237
FemalePopulation	*co-efficient*	(3.167)***	0.348	(1.489)
	p-value	0.000	0.709	0.319
	standard error	(0.614)	(0.926)	(1.489)
log(GDPpercapita)	*co-efficient*	(0.002)	0.003	0.006
	p-value	0.749	0.644	0.623
	standard error	(0.005)	(0.007)	0.011
WorkingAge	*co-efficient*	0.087	0.092	0.455
	p-value	0.682	0.773	0.375
	standard error	(0.210)	(0.317)	0.510
Note:	**p<0.1; **p<0.05; ***p<0.01*			

6.4 Discussion

The investigation into the level of financial inclusion in SSA revealed South Africa as the most financially inclusive country among the sample countries. The results are in line with the reports of The World Bank (2013) on the state of financial inclusion in South Africa compared to the rest of Africa. However, other researchers, such as Abdulmumin et al. (2019), only considered South Africa a low to middle-level inclusion economy. Asides from South Africa, the remaining countries did not differ too much in inclusion level; however, the level of financial inclusion is high in Namibia, Mauritius and Botswana, all members of the Southern Africa Development Community. The countries with low inclusion based on the study are Seychelles, Chad, and South Sudan. Another significant finding is the annual increase in the average level of financial inclusion in SSA for the study period.

The increase in the number of people that reside in urban cities can significantly increase the level of inclusion in SSA economies. While this potentially can mean further congestion of urban areas and the marginalisation of rural communities, the results align with previous research conducted by Cicchiello et al. (2021). The researchers recognised the negative impact of an increasing rural population on financial inclusion in a study carried out on 42 countries in Africa and Asia from 2010 to 2019. It is expected that rural communities have less access to financial services. Similarly, a positive relationship across all indices of financial inclusion could be seen between the number of adults within the working age range and financial inclusion.

The impact of electricity access on financial inclusion is significant to the financial inclusion domain in Africa. In the United Nations' Sustainable Development Goals report of 2021, SSA remained the region with the lowest level of electricity access in the world despite a 13% increase between 2010 and 2019 (United Nations, 2021). For every percentage increase the region makes to better the 46% recorded in 2019 to the global average of 90%, financial inclusion is expected to increase by 0.285%. The results also tilt the impact of electricity on the side of financial services usage rather than access. The advent of mobile banking and mobile money services drives the need for electricity access for the use of financial products and services.

The result of this study shows the positive effect of internet usage on financial access while there is a negative effect on usage and the overall level of financial inclusion. Increased exposure to the internet can help banks and other financial services better reach the population; however, the competition comes from novel financial products and services that might be difficult to capture in the financial inclusion index. A similar positive effect is recorded between internet usage and financial inclusion access in Latvia and Slovenia (Bayar, Gavriletea and Păun, 2021). On the other hand, Lenka and Barik (2018), who studied the relationship between financial inclusion and internet usage in South Asian Association for Regional Cooperation (SAARC) countries between 2004 and 2014, had contrary results with a 1% increase in internet use leading to 0.316% increase in the overall level of financial inclusion in the region.

There is a known gap in financial inclusivity between the male and female gender. However, the Sub-Saharan Africa region, alongside the Middle East and North Africa regions, are the areas that have regressed in breaching that gap between 2011 and 2017 (World Bank Group, 2021). The result of this study presents the fundamental challenge of an increased female population on financial inclusion. A 2.9% drop in financial access for every additional 1% of more females compared to men in SSA suggests restrictions and barriers to women accessing finance in the region. This result is similar in the adjusted model and even more negatively influenced in the model adjusted for outliers. The gap in access between men and women has been attributed to a low financial literacy rate, women using other accounts and a preference to use informal financial services (Fanta and Mutsonziwa, 2016). The findings also resonate with Adegbite and Machethe (2020), who confirmed the increasing financial inclusion gender gap caused by socioeconomic challenges limiting female access to formal financial services.

There is a positive but insignificant effect of economic growth, measured by GDP per capita in this study, on the level of financial inclusion. This is similar to the findings of Zhang et al. (2022) in their research based on 21 OECD countries. This is an interesting finding considering the significant positive impact of financial inclusion on economic growth evidenced by past literature (Emara and el Said, 2021; Younas, Qureshi and Al-Faryan, 2022). Therefore, it is expected that the focus and policy implementation drive targeted at the other determinants examined can lead to growth in financial inclusion, which in turn boosts economic growth and not vice versa.

6.5 Limitations

The research could only consider financial inclusion supply side data for 22 sub-Saharan African countries due to the unavailability of data. This represents less than 50% of the SSA member countries. Furthermore, the most comprehensive demand side database, the World Bank Global Findex database, is only triannual and equally contains many missing data entries for members of the SSA region. Also, the missing data points in the dataset for the selected countries required diverse imputation techniques such as mean substitution and probabilistic principal component analysis, which can potentially influence the analysis results.

7 Conclusion and Future Work

This study examined the impact of changing socioeconomic factors on financial inclusion with a focus on sub-Saharan Africa using fixed effect regression models. The urban population and electricity access rate had the most significant positive effect on financial inclusion. In contrast, unemployment and the female population ratio proved to have adverse impacts on the financial inclusion level in SSA. The financial inclusion index was computed using a two-stage robust principal component analysis to accommodate the presence of outliers and missing data entries imputed with probabilistic principal component analysis. The results suggest that policymakers can formulate policies geared at increasing access to financial products and services in rural areas to negate the impact of the increasing urban population. Electricity access remains a region-wide challenge. This study proves that significant efforts to increase electricity penetration and stability will yield economic benefits through financial inclusion. The results also make a case for the improvement of financial access to the female population for growth in financial inclusion as an increased female population had a positive effect on the usage of financial products. Despite the validity of the results, there was an inability to adequately represent many of the countries within the SSA region because of unavailable data. Further work should expand the scope of the study by carrying out a cross-country examination of the socioeconomic characteristics and financial inclusion through the lens of national income levels.

References

Abdul Karim, Z., Nizam, R., Law, S.H. and Hassan, M.K. (2021). 'Does financial inclusiveness affect economic growth? new evidence using a dynamic panel threshold regression', *Finance Research Letters*, p.102364. doi:10.1016/j.frl.2021.102364.

Abdulmumin, B.A., Etudaiye-Muhtar, O.F., Jimoh, A.T. and Sakariyahu, O.R. (2019). 'an investigation into the level of financial inclusion in Sub-Saharan Africa', *Scientific Annals of Economics and Business*, 66(1), pp.41–63. doi:10.2478/saeb-2019-0004.

Adegbite, O.O. and Machethe, C.L. (2020). 'Bridging the financial inclusion gender gap in smallholder agriculture in Nigeria: An untapped potential for sustainable development', *World Development*, 127, p.104755. doi:10.1016/j.worlddev.2019.104755.

Aït-Sahalia, Y. and Xiu, D. (2019). 'A Hausman test for the presence of market microstructure noise in high frequency data', *Journal of Econometrics*, 211(1), pp.176–205. doi:10.1016/j.jeconom.2018.12.013.

Alnabulsi, Z.H. and Salameh, R.S. (2022) 'The impact of launching the financial inclusion strategy on economic development', *Journal of Management Information and Decision Sciences*, 25(1), pp. 1–18.

Amponsah, M., Agbola, F.W. and Mahmood, A. (2021) 'The impact of informality on inclusive growth in Sub-Saharan Africa: Does financial inclusion matter?', *Journal of Policy Modeling*, 43(6), pp. 1259–1286. doi:10.1016/j.jpolmod.2021.03.009.

Baltagi, B.H. (2001). *Econometric analysis of panel data*. Uitgever: New York: John Wiley.

Banna, H., Kabir Hassan, M. and Rashid, M. (2021) 'Fintech-based financial inclusion and bank risk-taking: evidence from OIC countries', *Journal of International Financial Markets, Institutions and Money*, 75, p. 101447. https://doi.org/10.1016/j.intfin.2021.101447.

Bayar, Y., Gavriletea, M.D. and Păun, D. (2021) 'Impact of mobile phones and internet use on financial inclusion: empirical evidence from the EU post-communist countries', *Technological and Economic Development of Economy*, 27(3), pp. 722–741. doi:10.3846/TEDE.2021.14508.

Bekele, W.D. (2022). 'Determinants of financial inclusion: a comparative study of Kenya and Ethiopia', *Journal of African Business*, pp.1–19. doi:10.1080/15228916.2022.2078938.

Bell, A., Fairbrother, M. and Jones, K. (2018). 'Fixed and random effects models: making an informed choice. *Quality & Quantity'*, [online] 53(2), pp.1051–1074. doi:10.1007/s11135-018-0802-x.

Besong, S.E., Okanda, T.L. and Ndip, S.A. (2022) 'An empirical analysis of the impact of banking regulations on sustainable financial inclusion in the CEMAC region', *Economic Systems*, 46(1), p. 100935. t: doi:10.1016/j.ecosys.2021.100935.

Camara, N. and Tuesta, D. (2014). 'Measuring financial inclusion: a multidimensional index', *SSRN Electronic Journal*. doi:10.2139/ssrn.2634616.

Charfeddine, L. and Zaouali, S. (2022) 'The effects of financial inclusion and the business environment in spurring the creation of early-stage firms and supporting established firms', *Journal of Business Research*, 143, pp. 1–15. https://doi.org/10.1016/j.jbusres.2022.01.014.

Chipunza, K.J. and Fanta, A. (2021) 'Quality financial inclusion and its determinants in South Africa: evidence from survey data', *African Journal of Economic and Management Studies* [Preprint]. Doi:10.1108/AJEMS-06-2021-0290.

Cicchiello, A.F., Kazemikhasragh, A., Monferrá, S. and Girón, A. (2021). 'Financial inclusion and development in the least developed countries in Asia and Africa', *Journal of Innovation and Entrepreneurship*, 10(1). doi:10.1186/s13731-021-00190-4.

Danisman, G.O. and Tarazi, A. (2020) ‘Financial inclusion and bank stability: evidence from Europe’, *The European Journal of Finance*, 26(18), pp. 1842–1855. doi:10.1080/1351847X.2020.1782958.

Dar, S.S. and Sahu, S. (2022) ‘The effect of language on financial inclusion’, *Economic Modelling*, 106, p. 105693. Doi:10.1016/j.econmod.2021.105693.

Emara, N. and el Said, A. (2021) ‘Financial inclusion and economic growth: The role of governance in selected MENA countries’, *International Review of Economics and Finance*, 75, pp. 34–54. doi:10.1016/J.IREF.2021.03.014.

Fanta, A. and Mutsonziwa, K. (2016) ‘Gender and Financial Inclusion: Analysis of financial inclusion of women in the SADC region’. doi:10.13140/RG.2.1.1390.3605.

Fareed, Z., Rehman, M.A., Adebayo, T.S., Wang, Y., Ahmad, M. and Shahzad, F. (2022). ‘Financial inclusion and the environmental deterioration in Eurozone: The moderating role of innovation activity’, *Technology in Society*, 69, p.101961. doi:10.1016/j.techsoc.2022.101961.

Ghojogh, B., Ghodsi, A., Karray, F., Crowley, M., (2022) *Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: tutorial and survey*. doi:10.48550/arXiv.2101.00734

de Goeij, M.C.M., van Diepen, M., Jager, K.J., Tripepi, G., Zoccali, C. and Dekker, F.W. (2013). Multiple imputation: dealing with missing data. *Nephrology Dialysis Transplantation*, 28(10), pp.2415–2420. doi:10.1093/ndt/gft221.

Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P. and Acharya, A. (2019). ‘MICE vs PPCA: Missing data imputation in healthcare’, *Informatics in Medicine Unlocked*, 17, p.100275. doi:10.1016/j.imu.2019.100275.

Ifediora, C., Offor, K.O., Eze, E.F., Takon, S.M., Ageme, A.E., Ibe, G.I. and Onwumere, J.U.J. (2022). ‘Financial inclusion and its impact on economic growth: Empirical evidence from sub-Saharan Africa’, *Cogent Economics & Finance*, 10(1). doi:10.1080/23322039.2022.2060551.

Jarboui, S. (2021) ‘Renewable energies and operational and environmental efficiencies of the US oil and gas companies: A True Fixed Effect model’, *Energy Reports*, 7, pp. 8667–8676. doi:10.1016/J.EGYR.2021.04.032.

Kandari, P., Bahuguna, U. and Salgotra, A.K. (2021) ‘Socio-economic and demographic determinants of financial inclusion in underdeveloped regions: a case study in India’, *Journal of Asian Finance, Economics and Business*, 8(3), pp. 1045–1052. doi:10.13106/jafeb.2021.vol8.no3.1045.

Kebede, J., Naranpanawa, A. and Selvanathan, S. (2021) ‘Financial inclusion: measures and applications to Africa’, *Economic Analysis and Policy*, 70, pp. 365–379. doi:10.1016/j.eap.2021.03.008.

Khmous, D.F. and Besim, M. (2020) ‘Impact of Islamic banking share on financial inclusion: evidence from MENA’, *International Journal of Islamic and Middle Eastern Finance and Management*, 13(4), pp. 655–673. doi:10.1108/IMEFM-07-2019-0279.

Kim, D.W., Yu, J.S. and Hassan, M.K. (2018) ‘Financial inclusion and economic growth in OIC countries’, *Research in International Business and Finance*, 43, pp. 1–14. doi:10.1016/J.RIBAF.2017.07.178.

Kuada, J. (2019) ‘financial inclusion and the sustainable development goals’, in *Extending Financial Inclusion in Africa*. Elsevier, pp. 259–277. doi:10.1016/B978-0-12-814164-9.00012-8.

Kumar, V., Thrikawala, S. and Acharya, S. (2021) ‘Financial inclusion and bank profitability: Evidence from a developed market’, *Global Finance Journal*, p. 100609. doi:10.1016/j.gfj.2021.100609.

Lee, C.-C., Wang, C.-W. and Ho, S.-J. (2022) ‘Financial aid and financial inclusion: Does risk uncertainty matter?’, *Pacific-Basin Finance Journal*, 71, p. 101700. doi:10.1016/j.pacfin.2021.101700.

Lenka, S.K. (2021) ‘Relationship between financial inclusion and financial development in India: Is there any link?’, *Journal of Public Affairs*, p. e2722. doi:10.1002/PA.2722.

Lenka, S.K. and Barik, R. (2018) 'Has expansion of mobile phone and internet use spurred financial inclusion in the SAARC countries?', *Financial Innovation*, 4(1), p. 5. doi:10.1186/s40854-018-0089-x.

Li, P., Zhang, W., Lu, C., Zhang, R. and Li, X. (2022). 'Robust kernel principal component analysis with optimal mean', *Neural Networks*, 152, pp.347–352. doi:10.1016/j.neunet.2022.05.005.

Lu, W., Niu, G. and Zhou, Y. (2021) 'Individualism and financial inclusion', *Journal of Economic Behavior & Organization*, 183, pp. 268–288. doi:10.1016/j.jebo.2021.01.008.

Mahmoudi, M.R. *et al.* (2021) 'Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries', *Alexandria Engineering Journal*, 60(1), pp. 457–464. doi:10.1016/j.aej.2020.09.013.

Marcelin, I., Egbendewe, A.Y.G., Oloufade, D.K. and Sun, W. (2021). 'Financial inclusion, bank ownership, and economy performance: Evidence from developing countries', *Finance Research Letters*, p.102322. doi:10.1016/j.frl.2021.102322.

Mouna, A. and Jarboui, A. (2021) 'Understanding the link between government cashless policy, digital financial services and socio-demographic characteristics in the MENA countries', *International Journal of Sociology and Social Policy*. doi:10.1108/IJSSP-12-2020-0544.

Nguyen, T.T.H. (2021) 'Measuring financial inclusion: a composite FI index for the developing countries', *Journal of Economics and Development*, 23(1), pp. 77–99. doi:10.1108/JED-03-2020-0027.

Nuzzo, G. and Piermattei, S. (2020) 'Discussing measures of financial inclusion for the main euro area countries', *Social Indicators Research*, 148(3), pp. 765–786. doi:10.1007/s11205-019-02223-8.

Rojas Cama, F.A. and Emara, N. (2022) 'Financial inclusion and gross capital formation: A sectoral analysis approach for the MENA region and EMs', *International Review of Financial Analysis*, 79, p. 101993. doi:10.1016/j.irfa.2021.101993.

Sarma, M. (2008). Index of Financial Inclusion. Working Paper No. 215, New Delhi: Indian Council for Research on International Economic Relations

Sawadogo, R. and Semedo, G. (2021) 'Financial inclusion, income inequality, and institutions in sub-Saharan Africa: Identifying cross-country inequality regimes', *International Economics*, 167, pp. 15–28. doi:10.1016/j.inteco.2021.05.002.

Songwathana, K. (2018) 'The relationship between natural disaster and economic development: a panel data analysis', *Procedia Engineering*, 212, pp. 1068–1074. doi:10.1016/J.PROENG.2018.01.138.

The World Bank (2013) 'South Africa Economic Update: Focus on Financial Inclusion', *Africa Region Poverty Reduction & Economic Management*

United Nations (2021) 'The Sustainable Development Goals Report'.

Vaswani, N., Chi, Y. and Bouwmans, T. (2018) 'Rethinking PCA for modern data sets: theory, algorithms, and applications', *Proceedings of the IEEE*. Institute of Electrical and Electronics Engineers Inc., pp. 1274–1276. doi:10.1109/JPROC.2018.2853498.

Vo, D.H., Nguyen, N.T. and Thi-Hong Van, L. (2021) 'Financial inclusion and stability in the Asian region using bank-level data', *Borsa Istanbul Review*, 21(1), pp. 36–43. doi:10.1016/j.bir.2020.06.003.

Wang, R. and Luo, H. (Robin) (2021) 'How does financial inclusion affect bank stability in emerging economies?', *Emerging Markets Review*, p. 100876. doi:10.1016/j.ememar.2021.100876.

World Bank (2022) *Financial Inclusion Overview*. Available at: https://www.worldbank.org/en/topic/financialinclusion/overview [Accessed: 26 July 2022].

World Bank Group (2021) *The Drive for Financial Inclusion: Lessons of World Bank Group Experience*.

Xu, X. (2020) ‘Trust and financial inclusion: A cross-country study’, *Finance Research Letters*, 35, p. 101310. doi:10.1016/j.frl.2019.101310.

Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., & Lillard, J. W., Jr (2014). ‘A study of effects of multicollinearity in the multivariable analysis’, *International journal of applied science and technology*, *4*(5), 9–19.

Younas, Z.I., Qureshi, A. and Al-Faryan, M.A.S. (2022) ‘Financial inclusion, the shadow economy and economic growth in developing economies’, *Structural Change and Economic Dynamics*. doi:10.1016/J.STRUECO.2022.03.011.

Younes, K., Mouhtady, O., Chaouk, H., Obeid, E., Roufayel, R., Moghrabi, A. and Murshid, N. (2021). ‘The Application of Principal Component Analysis (PCA) for the Optimization of the Conditions of Fabrication of Electrospun Nanofibrous Membrane for Desalination and Ion Removal’, *Membranes*, 11(12), p.979. doi:10.3390/membranes11120979.

Zhang, P., Zhang, M., Zhou, Q. and Zaidi, S.A.H. (2022). The relationship among financial inclusion, non-performing loans, and economic growth: insights from OECD countries. *Frontiers in Psychology*, 13. doi:10.3389/fpsyg.2022.939426.