Configuration Manual

MSc Research Project
Fintech

Omotolani Opeyemi Uji
Student ID: 21200050

School of Computing
National College of Ireland

Supervisor: Victor Del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	..... Omotolani Opeyemi Uji .....
Student ID:	.....21200050.....
Programme:	.....Msc. Fintech..... **Year:** ...2022/2023.....
Module:	.....Msc. Research Project.....
Lecturer:	.....Victor Del Rosal.....
Submission Due Date:	.....August 14, 2023.....
Project Title:	.....Configuration Manual .....
Word Count:	.....1464..... **Page Count:** ...11.....

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):	

Configuration Manual

Omotolani Opeyemi Uji
21200050

1. Introduction

The configuration manual will provide details about the technical processes that were followed to be able to reproduce the analysis done to analyze the "*Impact of Economic growth and Financial stability on Financial inclusion in Sub-Saharan Africa*".

1.1 Hardware

- System type: MacBook Air, M1, 2020, 13.3 inch display
- Mac Operating System : Ventura 13.4
- Installed memory : 8.00 GB
- Computer name: MacBook Air

1.2 Software

R programming Language for Data Analytics Version 2023.03.0+386
Microsoft word to write the full report
Microsoft Excel to view and initially inspect the data for upload into R-studio

1.3 Data

Financial Inclusion – Data on financial Inclusion was downloaded from the IMF Financial Access Survey website. This was named "FAS" data.

Financial Stability - Data for financial stability was downloaded from the World Bank Global Financial development database. This was named "FINSTABDATA".

Economic Growth - Data for Economic Growth was downloaded from the World Bank World Development Indicators database. This was named "ECOGROWTH".

All the data were downloaded in CSV format.

https://www.worldbank.org/en/publication/gfdr/data/global-financial-development-database
https://databank.worldbank.org/source/world-development-indicators
https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C&sId=1460043522778

2. Data Analysis

2.1 Installing the packages for analysis

Amelia (version 1.8.1) - To visualize and treat missing data points

Stringr (version 1.5.0) – to handle "NA" values and smooth data flow between functions.

Dplyr (version 1.1.2) - to help in data manipulation

Psych (version 2.3.6) – used for validation tests such as Bartletts test and KMO test.

Car (version 3.1.1) – To check for multicollinearity in the dataset, the vif function was contained in the car package.

Stargazer (version 5.2.3) – To obtain the coefficients for the regression

plm (version2.6-3) - Linear models for panel data estimated using the lm function on transformed data.

2.2 Data Preprocessing and transformation

Preprocess the Financial Inclusion, Financial stability, and Economic growth data before they are merged into one dataset.

2.2.1 Financial Access Survey Dataset

- Importing the dataset -

```
FAS <- read.csv("FAS DATA NEW.csv", header = TRUE, na.strings = c(""))
```

- Selecting the indicators used in the study -

```
FASDATA1 <- FAS %>% select(Economy,
                           Year,
                           Number.of.deposit.accounts.with.commercial.banks.per.1.000.adults,
                           Number.of.registered.mobile.money.accounts.per.1.000.adults,
                           Number.of.commercial.bank.branches.per.100.000.adults,
                           Number.of.ATMs.per.100.000.adults,
                           Number.of.registered.mobile.money.agent.outlets.per.100.000.adults,
Number.of.mobile.money.transactions.during.the.reference.year..per.1.000.adults,
                           Outstanding.deposits.with.commercial.banks.REAL,
                           Outstanding.loans.from.commercial.banks.REAL)
```

- Change all the financial inclusion indicators to numeric

- Introduce a column known as CountNA to count the number of missing values in each row

FASDATA3$count_na <- rowSums(is.na(FASDATA3))

- Rename the columns because of their lengthiness

colnames(FASDATA3) <- c("Economy", "Year", "DepositAccts","MMAccts", "BankBranches", "ATMs", "MMAgents", "MMTransactions", "OutstandingDeposit", "OutstandingLoans", "CountNA")

- Removed countries with huge amount of missing values

2.2.2 Missing data treatment

- Check for missing values using the sapply and mismap function in r studio

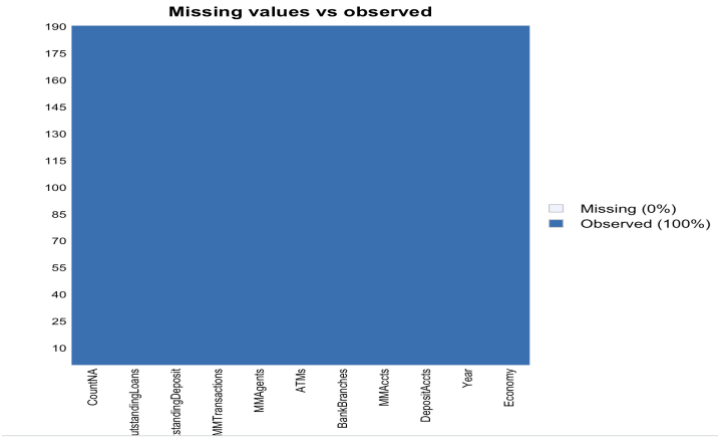
sapply(FASDATA30, function(x) sum(is.na(x)))

- Used Mean substitution to replace values that have preceding or following values in the same countries.

- Replace the countries that did not have mobile money in the years that are observed such as Mauritius, Gambia, Seychelles and Mauritiana with zero.

- Use the mismap to check for the missing data chart.

missmap(FASDATA30, main = "Missing values vs observed")



2.3 Financial Stability and Economic Growth Dataset

2.3.1 Financial Stability Dataset

- Importing the dataset

```
FINSTABDATA <- read.csv("FINANCIAL STABILITY DATA .csv", header = TRUE,
na.strings = c(""))
```

- Filter the dataset for the years and regions that are useful for the study

```
FINSTABDATA1 <- FINSTABDATA[FINSTABDATA$year > 2011, ]
FINSTABDATA2  <- subset(FINSTABDATA1, region == "Sub-Saharan Africa")
```

- Remove the countries that were already excluded from the Financial Inclusion Data

```
FINSTABDATA3 <- subset(FINSTABDATA2, country %in% c("Botswana", "Cameroon",
"Eswatini","Gambia.The", "Ghana", "Guinea", "Kenya", "Lesotho", "Liberia","Madagascar",
"Malawi", "Mauritius", "Namibia", "Rwanda","Seychelles", "South Africa",
"Uganda","Zambia","Zimbabwe"))
```

- Rename the financial stability indicators from s101, s102, s103, s106 for proper understanding of the data.

```
colnames(FINSTABDATA4) <- c("Economy", "Year", "Zscore", "NPL", "BankCap.Assets",
"LiquidAsset.Deposit", "count_NA")
```

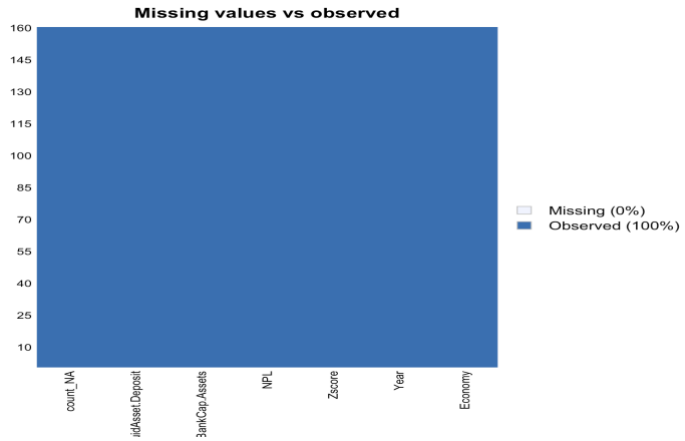
- Check for missing values with the sapply

```
sapply(FINSTABDATA6, function(x) sum(is.na(x)))
```

- Replace the missing values with Mean substitution

- Use the mismap to check for the missing data chart.

```
missmap(FINSTABDATA6, main = "Missing values vs observed")
```



2.3.2 Economic Growth Dataset

- Importing the dataset

```
ECOGROWTH <- read.csv("ECONOMIC GROWTH DATA .csv", header = TRUE,
na.strings = c(""))
```

- There was little or no preprocessing done to the Economic growth Dataset because it was downloaded based on the countries that were picked in the Financial Inclusion Data and Financial Stability.

2.4 Robust Principal Component Analysis

2.4.1 Check the financial Inclusion dataset for suitability in performing the PCA

- Check the correlation of the IMFData using the cor function in R

```
CorIMFFAS <- cor(IMFFASDATA[,3:10])
```

- Use the cortest.bartlett available in the psych package for bartlett's test in R studio

```
cortest.bartlett(CorIMFFAS,n = nrow(CorIMFFAS))
```

Output;

```
> cortest.bartlett(CorIMFFAS,n = nrow(CorIMFFAS))
$chisq
[1] 25.06471

$p.value
[1] 0.6243109

$df
[1] 28
```

- Perform the KMO test, an overall MSA greater than 5% signifies that PCA can be performed on the dataset;

```
KMO(CorIMFFAS)
```

```
> KMO(CorIMFFAS)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = CorIMFFAS)
Overall MSA =  0.7
MSA for each item =
      DepositAccts            MMAccts       BankBranches               ATMs
              0.77               0.80               0.60               0.72
         MMAgents     MMTransactions OutstandingDeposit   OutstandingLoans
              0.70               0.70               0.73               0.63
>
```

2.4.2 First Stage Robust Principal Component Analysis

- Creating each of the Penetration, Availability and Usage Indexes

Availability Dimension

- Use the **PcaCov** function in R on each of the dimensions

availabilitypc <- PcaCov(IMFFASDATA[, c(5:7)], scale = TRUE,center = TRUE)

The columns that contain indicators of the availability dimension.

$$\frac{\sum_{j,k=4}^{s} \lambda_j \, P_{kyc}}{\sum_{j,k=3}^{s} \lambda_j}$$

- It produces the principal component scores for each country per year P_{kyc}
- Obtain the eigen values for the dimension

```
> print(availability.var)
[1] 1.6379915 0.9289880 0.4330206
```

- The same process is repeated for each of the other dimensions (Usage and Penetration) to obtain the scores and eigen values.

2.4.3 Second Stage Robust Principal Component Analysis

- The results obtained from the first stage of the RCA **penetration, availability and usage index** were combined with the IMFDATA to create an additional dataframe.
- Use the PcaCov function to perform rcpa on the new dataframe

PCS.FII <- PcaCov(IMFFASDATA[, c(11:13)],scale = TRUE, center = TRUE)

Columns for the Availability, Usage and Penetration Index

- Use boxplot function to examine the presence of outliers.

- The Mean Statistics for each country to know which country was the most financially inclusive among the 16 countries and also to know the trends over the years considered.

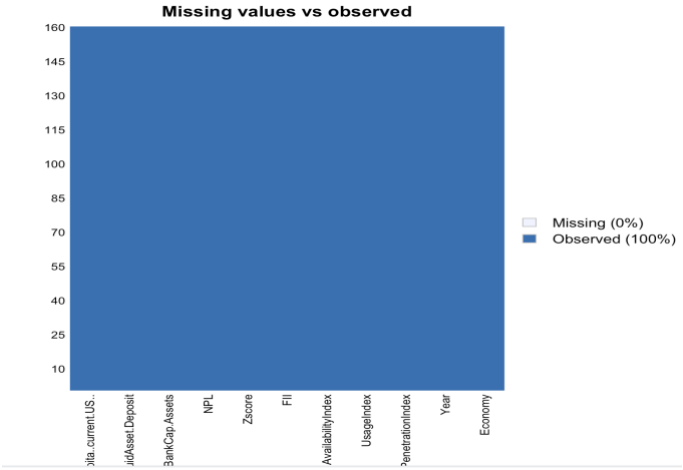
	FII.Group.1	FII.x	Usage.Group.1	Usage.x	Penetration.Group.1	Penetration.x	Availability.Group.1	Availability.x
1	Botswa	0.4101179	Botswa	0.3088570	Botswa	0.37959031	Botswa	0.3086508
2	Cameroon	0.2319445	Cameroon	0.2530175	Cameroon	0.09639321	Cameroon	0.2497818
3	Eswatini, Kingdom of	0.3186370	Eswatini, Kingdom of	0.2461968	Eswatini, Kingdom of	0.23981228	Eswatini, Kingdom of	0.3082119
4	Gha	0.3543523	Gha	0.2846501	Gha	0.29564929	Gha	0.2950625
5	Guinea	0.2164914	Guinea	0.2357017	Guinea	0.09232071	Guinea	0.2501487
6	Kenya	0.4281550	Kenya	0.3257069	Kenya	0.52039835	Kenya	0.2710319
7	Lesotho, Kingdom of	0.3127394	Lesotho, Kingdom of	0.2935966	Lesotho, Kingdom of	0.20771823	Lesotho, Kingdom of	0.2674210
8	Madagascar, Rep. of	0.2071837	Madagascar, Rep. of	0.2358452	Madagascar, Rep. of	0.08762275	Madagascar, Rep. of	0.2412507
9	Malawi	0.2285391	Malawi	0.2422736	Malawi	0.11823138	Malawi	0.2495390
10	Mauritius	0.8077622	Mauritius	0.6204628	Mauritius	0.66683740	Mauritius	0.3691366
11	mibia	0.5802717	mibia	0.4501315	mibia	0.42113283	mibia	0.3512918
12	Rwanda	0.3009907	Rwanda	0.2694378	Rwanda	0.21014028	Rwanda	0.2762387
13	Seychelles	0.7791787	Seychelles	0.3667817	Seychelles	0.68954528	Seychelles	0.5659658
14	South Africa	0.5102000	South Africa	0.3634196	South Africa	0.48485769	South Africa	0.3361915
15	Uganda	0.2778493	Uganda	0.2668322	Uganda	0.19679101	Uganda	0.2573516
16	Zambia	0.2737299	Zambia	0.2570651	Zambia	0.17033363	Zambia	0.2697315

2.5 Combined Dataset

- The IMFFASDATA was combined with the Financial Stability dataset and Economic growth Dataset;

FullDS <- cbind(Index, FSDATA1[,3:6], ECOGROWTH[3])

- Missing values were inspected with the mismap function in R



- The structure of all the variables in the dataset was changed to numeric except the Economy and year.

- The cor function in R was used to check the relationship among the variables.

3 Data Mining

3.2 Panel Data Analysis

We used variables to create a panel data that will be used for the analysis;

```
PanelData <- pdata.frame(FullDS, index = c("Economy" , "Year"))
```

3.2 Fixed Effects Model Estimation

Fixed effects estimation was carried out on the dependent variables on each of the independent variables using the plm function in R studio;

```
Model1 <- plm(log(UsageIndex) ~ Zscore + NPL + BankCap.Assets + LiquidAsset.Deposit + GDP.per.capita.current.US..,
        data = PanelData,
        model = "within")

summary(Model1)
stargazer(Model1, type = 'text')
```

NB: The results were analyzed with the summary and stargazer function

3.3 Random Effect Model Estimation

Random effects estimation was carried out on the dependent variables on each of the independent variables using the plm function in R studio

```
Model5 <- plm(log(UsageIndex) ~ Zscore + NPL + BankCap.Assets + LiquidAsset.Deposit + GDP.per.capita..current.US..,
        data = PanelData,
        model = "random")
summary(Model5)
stargazer(Model5, type = 'text')
```

3.4 VIF test for Multicollinearity

To check for the presence of multicollinearity, the VIF function was used from the car library in R;

```
install.packages('caret')
library(car)

Model9 <- lm(log(FII) ~ Zscore + NPL + BankCap.Assets + LiquidAsset.Deposit + GDP.per.capita..current.US..,
            data = PanelData)

vif(Model9)
1/vif(Model9)
```

3.5 Hausman Test

Comparing the result of the Fixed and Random Model Estimation

```
phtest(Model1, Model5) #alternative hypothesis
phtest(Model2, Model6)
phtest(Model3, Model7)
phtest(Model4, Model8)
```