NATIONAL BANK OF RWANDA
BANKI NKURU Y'U RWANDA

BNR ECONOMIC REVIEW
Vol. 17

February 2021

BNR ECONOMIC REVIEW
Vol. 17

February 2021

DISCLAIMER

FOREWORD

The bi-annual publication of BNR Economic Review intends to avail information to the public on economic matters, focusing on features and challenges of the Rwandan economy. This 17th volume of BNR Economic Review consists of four research articles touching on topical issues related to financial stability as well as price stability in Rwanda. The papers aim to provide concrete evidence-based analyses and policy recommendations that can help to improve the effectiveness of financial stability and monetary policies in Rwanda.

The first two articles in this volume address two topical issues in macro and micro-prudential policy debates, namely systemic risk and non-performing loans. The first paper explores an important feature of systemic risk related to similarities in banks' asset portfolios. The paper also investigates whether revealed portfolio similarity could contribute to the build-up of systemic risks. After exploring how the assets side of the balance sheet and lending portfolios evolved over time (2016-2019), the findings point out a relatively stable cosign similarity index of banks' portfolios. The index highlights four clusters of similar banks in three consecutive years (2017-2019). The first cluster encompasses banks that finance most of the sectors at an average level, except in agriculture where they exhibit higher levels of exposure. The second cluster consists of the second-lowest lenders to all the economic sectors. The third cluster comprises larger banks with portfolios covering most of the sectors, except in agriculture where they have the second-largest exposure. The fourth cluster consists of smaller banks that have the lowest exposure to all sectors except trade and mining. These findings highlight the importance of banks' portfolio similarity and clustering analyses for early detection of systemic risk exposure of the Rwandan banking system.

Cognizant of the increasing stock of non-performing loans (NPLs) among manufacturing sector credit portfolios of banks and its negative implications for the stability of the financial system in Rwanda, the second study investigates firm-level determinants of loan delinquency among manufacturing firms using a unique dataset compiled in 2018 from a survey on access to finance conducted among 122 manufacturing firms in Rwanda. The study uses various financial ratios to examine whether the surge in NPL was a result of financial distress and/or a moral hazard effect. Analysis of survey data suggests that, by 2018, about 32 percent of borrowers had histories of irregular or overdue repayment, most of whom self-reported challenges related to cash flow and decreased revenue. The analysis further shows higher incidences of credit-delinquency among sole-proprietors and smaller and younger firms. The results of a logit model reveal that none of the financial ratios for moral hazard are significant predictors of credit delinquency. In line with the theoretical expectation of the financial distress hypothesis, the results suggest that the odds of credit delinquency significantly decrease with profitability and liquidity ratios. Based on these results, the study outlines major implications in terms of banking regulation and financial stability policy in Rwanda.

The remainder of this volume comprises two studies that focus on the price stability objective of the National Bank of Rwanda (NBR). Following the adoption of a price-based monetary policy framework, the third study in this volume sets out to measure the magnitude of the economic cost associated with the medium-term inflation benchmark of 5±3 percent. Using quarterly observations from 2006:3 to 2019:4, the authors estimate two alternative money-demand function specifications, namely the log-log and semi-log models. The Dynamic Ordinary Least Squares (DOLS) and autoregressive distributed lad (ARDL) procedures are used to obtain cointegrating regressions for both specifications. Various diagnostic tests suggest that, compared to the semi-log specification, the double log demand function performs better on the Rwanda data. Estimates of the double-log specification from the ARDL procedure turn out to have more predictive power and suggest a welfare cost ranging from 0.5077 to 0.8537 percent of annual real GDP over the medium-term inflation benchmark band of 5±3 percent. Compared to other economies in the region (such as Kenya), these estimates are relatively higher, which supports the case for continued pursuance of price stability as the primary goal for monetary policy in Rwanda.

The transition to a price-based monetary policy framework in Rwanda has been the hallmark of the ongoing modernization of monetary policy by NBR. The forward-looking framework is supported by a Forecasting and Policy Analysis System (FPAS) that streamlines macroeconomic forecasting into monetary policy decision-making. The fourth article documents the econometric models used by the NBR for short-term inflation forecasting. The forecasts constitute key inputs in the quarterly projection model, which is the core of the FPAS. The paper shows that multivariate models outperform benchmark models and that, the forecasting combination reduces forecast errors compared to the individual model forecasts. These findings support the current use of combined multivariate models for short-term inflation forecasting to capture more information on economic dynamics.

Comments and questions can be sent to the Director of Research on phitayezu@bnr.rw, KN 6 Avenue 4, P.O Box 531 Kigali-Rwanda.

RWANGOMBWA John
Governor

TABLE OF CONTENTS

ASSESSING THE SYSTEMIC IMPORTANCE OF BANKS IN RWANDA USING PORTFOLIO SIMILARITY AND CLUSTERING METHODS

Dominique Ntirushwamaboko[1]
Patrick Mugenzi[2]
Christian Nyalihama[3]

1. Manager, Data science Division, Statistics Department , National Bank of Rwanda
2. Economist, Research Department, National Bank of Rwanda
3. Senior Principal Economist, Research Department, National Bank of Rwanda

ABSTRACT

The paper assesses the similarity among Rwandan banks, especially looking at how the assets side and lending portfolios have been evolving and their implications on systemic risks in the Rwandan banking system. The aim was to gauge a systemic risk that might originate from a cluster (s) of small banks, which is not well captured by traditional means of using the size or interconnectedness in network analysis. We used a variety of empirical approaches to tackle this aspect in the context of Rwanda with data from 2016 to 2019. Our key findings suggest that the general measure of the portfolio similarity between individual banks is quite stable over time and driven predominantly by big banks. Conversely, we noted that some medium-sized banks have been consistently similar in terms of the loan portfolio and associated risks in the last four years, hence they can be exposed to common risks with impactful consequence, as the clusters are more sizeable than individual banks.

Keywords: Systemic risk, Banks' portfolios, Cluster Analysis
JEL classification: G01, G11, C38

1. INTRODUCTION

Assessing linkages and complexity in financial systems has gained more interest in literature and policymaking, especially in the aftermath of the 2008 Global financial crisis (henceforth GFC). The existence of linkages and feedback loops is seen as leading to a high propensity of bank failure(s), which would affect the whole banking system and the real economy (Lux, 2016; Krause & Giansante, 2012).

The occurrence of bank failure(s) does not necessarily lead to similar consequences on financial system stability and the real economy. Some bank failure(s) are very consequential as the affected bank(s) can be systematically important due to different reasons. According to IMF, BIS & FSB (2009), systemic importance is difficult to define. Still, in practice, a financial institution can be established as systemic if its failure or malfunction causes general distress in the financial system, either as a straight impact or as a cause for wider contagion. The systemic importance of a bank(s) implies a high propensity that failure can impend the stability of the whole financial system. In the 2008 GFC, the failure of some systemically important banks spills over the entire financial systems with dire consequences on the latter and the real economy (Brechler, et al., 2014).

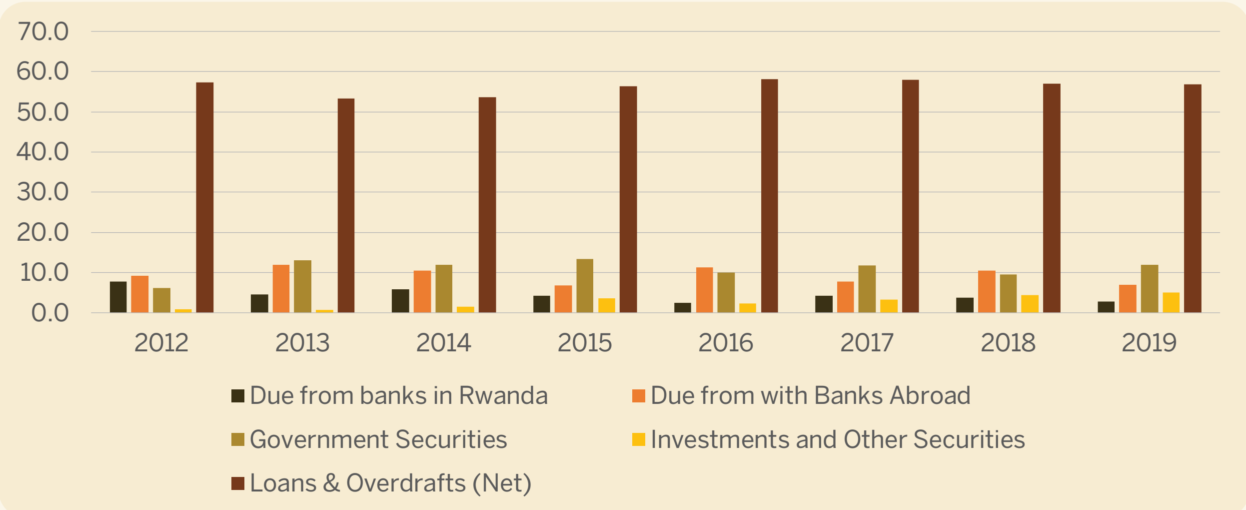
The literature on assessing the systemic importance of institution(s) in the banking system has mostly focused on direct channels. Direct channels involve direct contractual obligations between banks, for instance when a given bank has borrowed from another bank on the interbank market. These direct channels imply that a bank failure rapidly spillovers to the whole financial system via linkages, mostly in the form of financial exposure between banks as the failing bank will not be able to honor its contractual obligation. By domino effect, this will spread over the whole banking system. The financial exposure may be exacerbated by a subsequent negative impact on the real economy and existing feedback loops with the financial sector.

Direct channels often consider three criteria, namely, substitutability, size, and interconnectedness. Size in this situation denotes the volume of financial services and products provided by a certain bank in the banking system. The extent to which other banks can provide the same services is what is defined as substitutability. Lastly, interconnectedness refers to the degree of interdependence between banks or the extent to which an individual bank is linked to other banks (e.g. exposure on the interbank market which may cause bank failures when a bank is unable to pay what it owes to other banks (s)) in the financial system (IMF, BIS and FSB, 2009). Despite numerous challenges in measuring systemic importance, recent literature such as Leon et al. (2015) and Leon et al. (2018), etc., focused on interconnectedness and used network analysis methods to assess banks' systemic importance via centrality in money markets and payment systems.

However, another strand of recent literature (e.g., Brechler, et al., 2014; Leon, 2017) considers additional features of the banking system by looking at similarity/homogeneity in banks' portfolios as a possible source of systemic risks. The similarity of banks' portfolios (i.e. when banks have invested in similar assets such as loans per sector, government securities, etc.) can be across the whole financial system or concentrated in different clusters within the banking system (Brechler, et al., 2014). The gist behind this is that similarity of balance sheets makes exposure to common economic shocks more likely and intense (Leon, 2017), and this may have substantial implications on financial stability as losses in those financial institutions may be highly correlated. Indeed, common asset holdings, that is when banks invest more or less in the same assets, have been the main factor of contagion during the 2008 GFC (Caccioli, et al., 2014). The issue with common assets holding is that when one bank fails and its failure is followed by assets fire sales, the negative impact on prices of those assets will lead to failures of other banks as the latter have the same assets in their portfolio.

In Rwanda, the banking system experienced significant mutations for the last two decades. Its size has been consistently expanding along with the real sector. The number of banks has tripled from five banks in early 2000 to 16 banks currently, and the balance sheet sizes have grown following mostly increasing volume of loans to finance economic activities. The level of concentration in terms of assets, loans, and deposits has also been declining, although at a slower pace in recent years. Despite recent developments in the interbank markets, the share of the amount due to other domestic banks in total assets is still relatively low.

Figure 1: Evolution of share of main assets categories in total assets of the banking system



Source: National Bank of Rwanda (NBR) & Authors' computation

In a landscape of the banking industry with a low level of concentration and exposure to other banking institutions, assessment of systemic importance via size or interconnectedness in interbank markets may suggest a situation with lower systemic risks. Nevertheless, in the context of less developed financial markets, banks are likely to hold similar investment and risk management strategies, and that homogeneity can contribute to systemic risks (Leon, 2017). In Rwanda, lending has been the primary income-earning activity with the highest share on all bank balance sheets. Also, banks' lending has consistently been channeled to some key sectors such as mortgage, trade, and transport, and this common exposure may predispose the banking industry to systemic risks once some sectorial shocks erupted. From that perspective, one would wonder how similarity in terms of the structure of financial statements across banks has evolved in Rwanda and its implications on the assessment of systemic risks.

This study aims to assess how similarity among Rwandan banks, especially on the lending portfolios, has been evolving and implications on systemic risks in the Rwandan banking system. This study borrows from Brechler et al. (2014) in their study for the case of the Czech Republic. It applies a similar methodology to measure similarity and identify the existence of clusters of similar banks, which may turn out to be systemic.

Similarity and clustering methods allow us to gauge systemic risks stemming from a cluster(s) of small banks, which is not well captured by traditional ways of using the level of interconnectedness in network analysis. Because on one side, similarity and clustering methods help to assess vulnerabilities that do not directly come from a breach of contract by another bank, and on the other side, they can help to discover the importance of small banks which usually have lower centrality in network analysis.

Given that in Rwanda, contractual obligations notably interbank transactions and other due from one bank to another bank have remained relatively low over time, it is worthy to look at indirect channels, to evaluate whether homogeneity in exposure to a given sector/market/debtors may be a source of systemic vulnerability. Therefore, this study proposes an additional systemic risk-monitoring tool to complement existing ones based on direct channels (size, concentration, stress test and network analysis via cross exposure in interbank markets). It will inform financial regulation and ensure appropriate monitoring of systemic risks build up and timely actions to protect the financial system against threats coming from common exposure or clusters.

The structure of this study is as follows: the next chapter reviews the empirical literature on measuring systemic importance within the banking system. Chapter 3 overviews the evolution of the banking system in Rwanda. Chapter 4 explains the methodology used. Chapter 5 details empirical results and Chapter 6 concludes.

2. EMPIRICAL LITERATURE

Systemic risk is at the center of all financial stability practices, and the recent worldwide financial crisis has awakened interest in the subject. The literature on measuring systemic importance has been growing with different approaches, mostly in line with various forms through which systemic importance may materialize. ECB(2010) distinguished three primary types namely contagion risks from of individual firms (using measures such as expected shortfall, conditional value at risk), the common exposure to adverse macroeconomic shock (using stress testing, etc.), and the dangers from widespread financial imbalances (e.g., credit cycles, leverage, maturity mismatch, etc.). In addition to these, other new approaches emerged in the last decade, including the network analysis to understand the structure and interconnectedness in the banking system (Brechler, et al., 2014). Of recent, using measures of similarity/homogeneity in banks portfolio and clustering have surfaced as another approach to measure systemic importance.

For the first approach using individual risk to individual institutions, one example is a study by Drehmann and Tarashev (2011) who assessed systemic importance in the global banking market using Shapley value. These were derived as portions of system-wide risk that are ascribed to separate institutions. They used expected shortfall as a measure of systemic risks arising from the systemic event, and use both the participation approach (where the increment of systemic risks is attributed only to an interbank lender) and the contribution approach which measures risks generated by the bank itself as well as how it contributes to the uncertainty in other subsystems. Using data of 20 large international banks, they show that both approaches can lead to various actions of systemic importance, and regulators should pay attention to that.

As highlighted in previous sections, the network analysis, notably in the interbank market, is also another approach used to gauge systemic risks. As emphasized by Gai et al. (2011); contagion, complexity, and concentration can play a significant role in financial fragility aftershock. Several studies (e.g., Sheldon & Maurer, 1998; Espinosa-Vega & Sole, 2010; Minoiu & Reyes, 2011; Leon, et al., 2015; Leon, et al., 2018 ; Csonto, et al., 2018 ; Chretien, et al., 2020), adopted the network analysis, to identify systemically important financial institutions. For country case, Leon et al. (2018) studied on the structure of the Colombian interbank market analyze interconnectedness and hierarchy looking at linkages between market players and how central bank liquidity is allocated throughout the market. They used a measure of centrality and identified some key players who have a pivotal role as a super spreader of central bank liquidity on the interbank market. They also show that size is a crucial determinant for banks to be super spreader as larger banks were found to be central and interconnected in the network, which could intensify contagion.

Leon et al. (2015) had also used network analysis to assess banks' centrality on the money market and the significant value payment system as one of the metrics in determining the systemic importance in the Colombian banking system, in addition to size and non-substitutability. Results suggest a skewed distribution on both the money market and payment system where few financial institutions are vital while a large number of others are of less importance in the network.

Chretien et al. (2020) took an encompassing view on issues of interconnectedness in the financial system (including banks, insurance, and mutual funds in France) by considering two contagion channels. One channel through prices of assets, which are commonly held by several financial institutions, the second channel is loss due to defaults by counterparts, which lead to a cascade of default. Their results show that the propagation of shock is through the first channel is more critical than in the second channel. The study also highlighted the difference in role played by banks, insurance, and investment funds in contagion.

In the past, Sheldon and Maurer (1998) had also used to the network analysis to study the systemic risk in Switzerland by analyzing the probability of a bank to fail in any period considering the route centered on the interbank loan structure. They also found the likelihood that a shock in one bank will take, and the effect of this shock on the creditworthiness of other banks connected via loans to that defaulting bank. The data used are the interbank transactions limited to short-term (0-3 month) interbank time loans and deposits. Their study revealed that the structure of interbank loans had a higher likelihood of bank insolvency in any given year is quite high. On the contrary, the probability of a bank failure spreading via the financial system through the network of interbank loans were found low.

Regarding cross-country studies, Csonto et al. (2018) assessed the banks or deposit-taking corporations' vulnerability to shocks via the interbank linkage of locally incorporated banks in the Eastern Caribbean Currency Union (ECCU) financial system. They used the balance-sheet network analysis technique in Espinosa-Vega and Sole (2010), and the indicators employed were the number of bank failures, the average percent of asset loss in other banks (failed capital), number of contagion rounds, and the percent of capital loss in the system (index of contagion). They looked at how the individual bank's vulnerability is determined, to which extent is open to the failure of other banks in the system, and applied the funding shock and the credit shock, and as a simulation of shocks. In the case of credit shock, a bank is supposed to suffer a loss on the asset side of its balance sheet. The shock and the credit shock happens when other banks in the network default. As the bank records its claims to other banks in its assets, the value of its claim declines and hence loss on the asset side. Under a funding shock, it is expected that there is an abrupt drawing of funding from every bank in the network system, thus causing a shock from the liability side.

The results in an abrupt gap in funding that, in return, affects the balance sheet of the bank if it fails to obtain an alternative funding source to substitute it. The results of their study revealed the systemic importance of some regional banks in the ECCU region and the vulnerabilities of local banks.

Previously, Minoiu and Reyes (2011) study had also used a network analysis technique to assess dynamics in the global banking network between 1978 and 2009. They used three metrics of centrality, connectivity, and clustering with data on cross border bank lending from 184 countries. Their results show that the topology of the global banking network has been volatile and was affected by the cycle of capital flows, banking, and sovereign debt crises. In particular, the 2008 GFC negatively affected connectivity in the global banking network.

There has been a substantial development in the literature works centering on the causes and sources of the systemic importance of banks or financial institutions, but very little in practice frequently make simple the problem of size and contagion owing to interbank market linkages (Brechler, et al., 2014). Some recent papers (e.g., Brechler et al., 2014; Cai et al., 2018; Leon, 2017, etc.) adopted measuring similarities/homogeneity amongst financial services providing institutions and the existence of clusters within the financial system as a new approach to gauge systemic importance in financial systems.

Leon (2017) studied homogeneity (defined as lack of diversity due to some forms of uniform diversification) in the Colombian banking system by measuring their similarity in terms of lending, funding, and investment portfolios.

His study delved deeper into bank granular data and used agglomerative clustering with machine learning techniques. Besides, similarity and clusters were identified using the Euclidian distance between banks and the Ward linkage method. The result revealed the existence of some degree of similarity in the Colombian banking system, especially with regards to lending and funding, while investment portfolios are relatively less homogenous. Also, shreds of evidence suggested that homogeneity was more substantial in the largest banks.

In a precedent study, Cai et al. (2018) had also used Euclidian distance as a measure of similarity, focusing on the US syndicated loans market from 1988 to 2011. They also calculated the aggregate interconnectedness index, which indicated an increase in the interconnectedness between 1989 and 1994 and a significant decline in 2008 and 2009 before a pick up afterwards. Evidence also suggested that banks tended to concentrate syndicate lenders and that bank size, the level of diversification and specialization are positively correlated with its interconnectedness and diversifications matters more.

Lastly, Brechler et al. (2014) used similarity and clustering methods to examine banks' systemic importance in the Czech Republic. Their approach was different as they used the Cosine similarity function to measure the similarity between bank assets and loan performance instead of Euclidian distance, which is a measure of dissimilarity.

Evidence from data spanning from 2002 to 2013 suggests that overall similarity has generally been stable and not excessively high. Nevertheless, the similarity was high in big and deep-rooted commercial banks, and some clusters of small banks, which could turn out to be systemically important were identified. Aldasoro and Alves (2016) also used cosine similarity in their study of similarity between layers on the multiplex interbank network of large European banks.

Regarding clustering methods, a study by Tabak et al. (2011) on systemic risk in Brazil used a directed clustering coefficient in a complex banking network. The authors employed daily data on loans collected from financial services providing institutions within the Brazilian financial system for all deposit-taking corporations that have exposures in the interbank market system from January 2004 to November 2007. Their results showed that the directed clustering coefficients are an indicator of a systemic risk and that these indicators vary over the financial institutions, and they are negatively associated with the change in interest rate. They concluded that banks change their risk exposure with changes in interest rates. Generally, systemic risk in Brazil was found to be very limited.

Another study conducted by Dardac and Boitan (2009) assessed homogeneity in terms of risk profile and profitability in Romania banking industry applying cluster analysis and identified large and complex banking groups as a potential source of systemic risk in Romania.
They used financial intermediation data from the period of 2004 to 2006 and applied the clustering techniques for each of the three years, taking into account single linkage, complete linkage, and centroid clustering for computing distance functions.

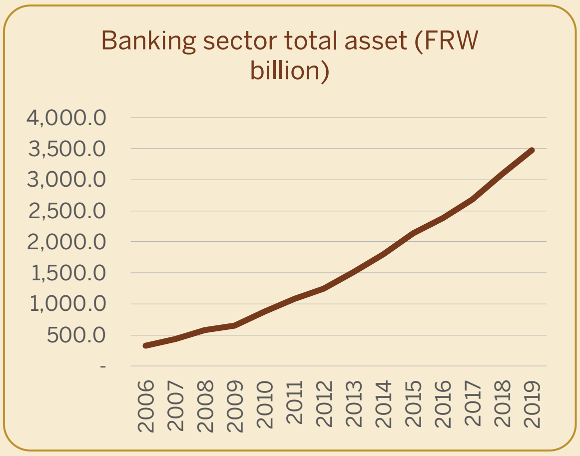
The present study implements the similarity and cluster analysis to examine systemic risk based on similar features of the credit exposure and balance sheet in the Rwandan banking system. It heavily borrows from Brechler et al. (2014) and Dardac & Boitan, (2009) to examine the Rwandan case; the empirical strategy is detailed in the next sections.

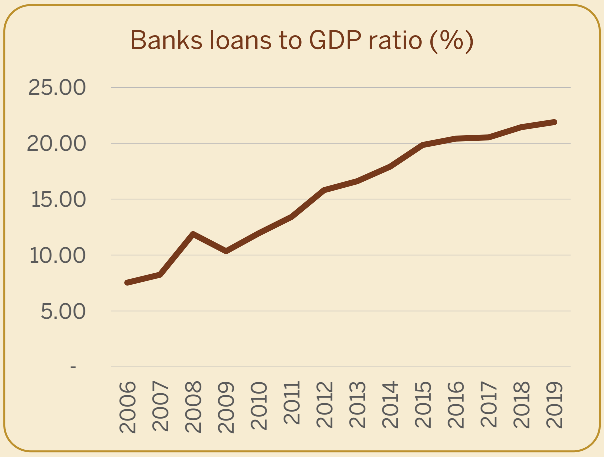
3. OVERVIEW OF THE BANKING SECTOR IN RWANDA

The banking system in Rwanda underwent significant progress in the last two decades in various areas. The number of banks grew from six in 2006 to sixteen in 2020, including several foreign banks entering the market. Expansion in terms of banking institution size is also noticeable in line with growth in the real sector. Increasing the banking sector lending to the economy has also played a role in the process. (Nyalihama & Kamanzi, 2019).

Figures 2 and 3 below illustrate the rising role of the banking industry of the Rwandan economy. They express that the total assets of the banking industry have grown over four times in the last ten years, driven particularly by lending activities. Loans to GDP ratio doubled during the same period from 11.9% in 2010 to 21.9% in 2019. Nevertheless, this level is still relatively low compared to emerging economies.

Figure 2: Banking sector total asset (FRW bn) **Figure 3:** Banks loans to GDP ratio (%)





Source: NBR and Authors' calculation

Financial market development is still at an early stage in Rwanda, and financial intermediation is generally deposit taking and lending by commercial banks where lending is the primary income-generating activity (see Figure 6). Deposits are the primary source of funds (see Figure 5). For the last seven years, loans share in total assets has remained steady, around 57%, followed by Government securities with a share of about 10% (see Figure 4). On the other side, deposits remain the primary source of funds with the share in total liabilities is currently at 63% down from 68.4% in 2014 (see Figure 5). Funds from other banks have been gradually expanding along with the deepening of interbank markets.

On deposits, the bulk of them is highly liquid demand deposits. By the end of 2019, the share of demand deposits stood at 43% against 34% for time deposits. On the type of depositors, Households share is still the largest, although dwindling amid the expansion of institutional deposits share. Despite these improvements, the banking sector still over rely on short-term funds, and this poses a challenge for maturity transformation.

The leverage ratio has remained high, and the capital adequacy ratio remained above 20% at the industry level well above the minimum required of 15%. Banking sector profitability has improved in the last two years, as shown by the return on equity for the industry (see Figure 7). However, it has been volatile in the past and mostly driven by well-established banks. Moreover, efficiency gains and improvement in asset quality, especially loans, as shown in Figure eight below, have contributed to this observed gain in banking sector profitability.

Figure 4: Banking sector assets decomposition

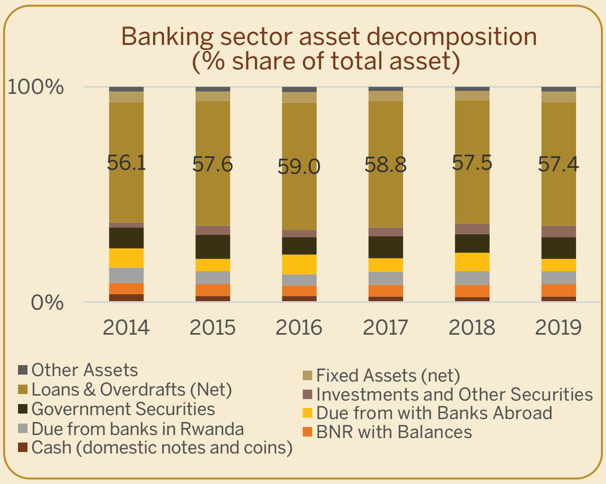


Figure 5: Banking sector source of funds

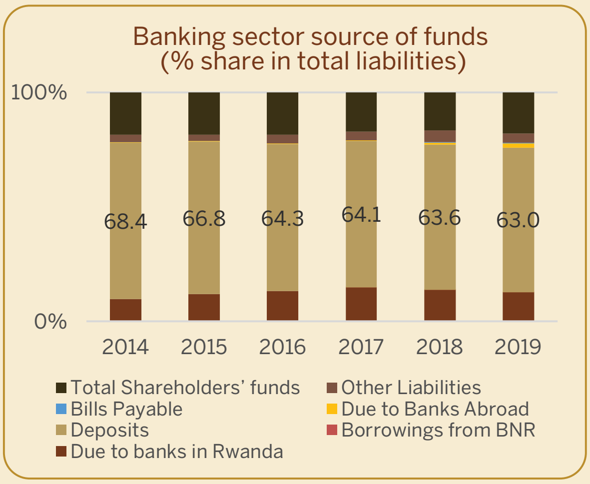


Figure 6: Banking sector source of income

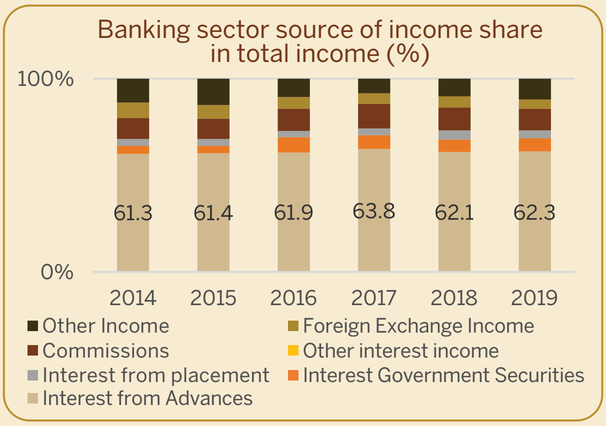
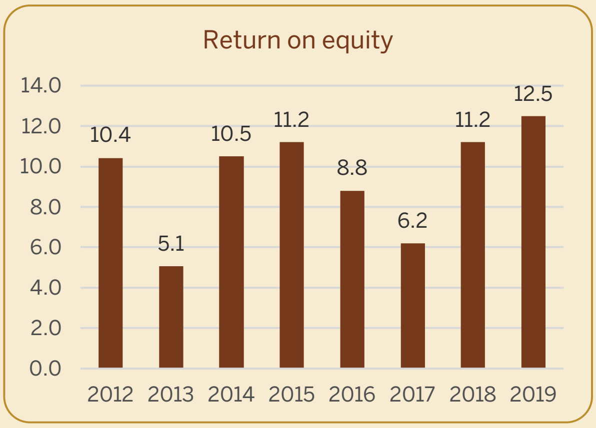


Figure 7: Banking sector return on equity



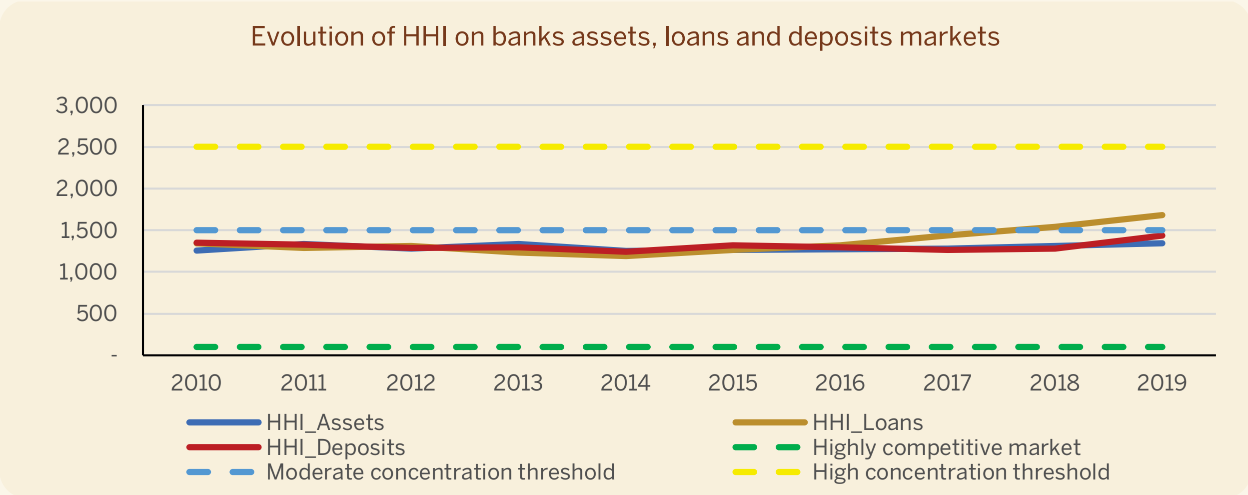
Source: NBR, Authors calculation

Figure 6: Non-performing loans to gross loans (%)

Source: NBR, Authors calculation

Overall, banking sector development in Rwanda is evident, and indicators of capital adequacy, liquidity, assets quality, and profitability have remained generally sound. Nevertheless, as shown in the previous sections, the source funds and assets remain concentrated in deposits and loans, respectively, which may turn out to be a vulnerability at the bank level. Besides, at the industry level, some progress recorded in previous years on assets, loans, and deposits market concentration has been waning for the last two years, as shown by Figure nine below. In particular, HHI for the loan market suggests that currently, the loan market is moderately concentrated.

Figure 6: Non-performing loans to gross loans (%)



Source: NBR and Authors' calculation

Even though banking sector in Rwanda heavily relies on short-term resources, the recent stress test on liquidity coverage ratios suggested that the banking sector in Rwanda was resilient to a sudden withdrawal of 10 percent of institutional investors' deposits and a simultaneous dry of 5 percent inflows from loans repayments (NBR, 2019)

4. METHODOLOGY

In the subsequent section, we describe steps undertaken to assess systemic importance in the banking system using similarity and clustering methods. As mentioned in the previous sections, this study follows the methodology used by Brechler et al. (2014) for the case of the Czech Republic.

In the beginning, we designed a measure of banks portfolios' similarity, which considers the similarity of categories of assets between individual banks. This measure helps to visualize the risk of banking system fragility, which links up with contagion risk; however, the system remains stable and unwavering with the absence of adverse shocks.

The analysis of common risks could thus be vividly crucial in understanding the vulnerability of the banking system, which may transform into financial system instability. Specifically, the similarity of banks' portfolios can be analyzed in different facets, namely the typical risk profile and the whole banking system without account for assets risk. In Rwanda, as most banks are in the traditional commercial banking business, the main item on their balance sheet is loans as discussed in the previous sections. For this reason, we also investigate similarity in Rwandan banks' loan portfolio and their risk profile. Similar to Brechler et al. (2014) , considering loans risks per sector help to better gauge systemic risks as loan portfolio similarity and the exposure to the same risks would exacerbate vulnerabilities.

Lastly, measures of similarity of the balance sheet, loan portfolio, and risks adjusted loan portfolio will help to identify clusters of very similar banks in Rwanda and how those clusters dynamics overtime. Because the Rwandan banking system includes both bigger and smaller banks in terms of assets, identifying clusters could help to see whether a cluster of smaller banks can turn out to be systemic if taken as a group.

Although banks can be characterized by the structure of the asset side of the balance sheet, its liquidity conditions, leverage conditions, among others, this study focuses on the asset side of 16 Rwandan banks' balance sheets notably banks loans portfolio given the importance of bank lending as the main source of external funds for corporate sector and households in Rwanda. The methodology described below could be applied on the liability side as well and this may be undertaken in future studies.

Specifically, in the first place, we consider the assets side components and narrow down to one asset category, namely loans portfolio thereafter. As in Brechler et al. (2014), let's consider a vector a = {a1...ak}, where k stands for data and, denotes the asset portfolio, characterized by the combined gross nominal value of each asset category $i \in 1,..,k$. In our case, assets categories are cash, balance with the central bank, due from other banks in Rwanda, due from other banks abroad, Government securities, investment and other securities, loans, fixed assets, and other assets. Regarding loans, granularity is according to the economic sector in which loans are directed to.
Following Brechler et al. (2014), we measure the similarity between the portfolios of any two banks (e.g., x and y) utilizing a cosine similarity function. Cosine similarity has several advantages, including the fact that it is a scale-independent measure and is bounded to $\{a, b \in R\} \to [-1, 1]$ by definition. Furthermore, given that balance-sheet assets only take non-negative values, we further find $\{a, b \in R+\} \to [0, 1]$, 0 for orthogonal vectors, which is the complete dissimilarity and 1 for identically oriented vectors which is the completely identical portfolio composition. The cosine similarity between two vectors (a and b) is defined as the cosine of the angle between the vectors:

$$\text{similarity (a,b)} = \cos(\theta) = \frac{a.b}{\|a\|\|b\|} = \frac{\sum a_i b_i}{\sqrt{\sum a_i^2 \times \sum b_i^2}} \qquad (3.1)$$

The above formula is used to measure similarity in banks' asset categories and loans per sector. In the second phase, loans per sector are adjusted with a measure of credit risks (details in the next sections). In line with Brechler et al. (2014), the weighted cosine similarity is defined as follows:

$$\text{similarity}_{wt}(a,b,w)= \frac{\sum w_i a_i b_i}{\sqrt{\sum w_i\, a_i^2 * \sum w_i b_i^2}} \tag{3.2}$$

As indicated above, a measure of credit risk is included as a weight in equation 3.2. Following Brechler et al. (2014), the level of credit risk associated with a certain asset category at time t, μt, is measured by the combined value of NPLs ratio per sector for the whole banking industry in a period under consideration. We also consider the coefficient of variation of those NPLs ratios per sector to gauge the overall similarity of NPL ratios across banks, and the risk weights are derived as the level of NPLs and the inverse of the coefficient of variation. The latter is inverted because the higher coefficient of variation indicates dissimilarity instead of similarity. The formula used is as follows:

$$W_t = \mu_t * V_t^{-1} \tag{3.3}$$

$$V_t = s_t / \mu_t \tag{3.4}$$

Where μ_t NPL ratio per sector at time t and s_t is standard deviation at time t. If we compute a (di) similarity measure (3.1) for each pair of n banks, we get an n×n symmetric similarity matrix S = {si, j} that denotes an essential input for deriving several characteristics of the banking system. Form similarity matrix S, we derive a measure of the overall similarity in the Rwandan banking system during the sample period. Secondly, in addition to matrix S, the matrix with risk-adjusted loans also help to identify clusters of banks exposed to joint risks, which can eventually turn out to be systematic.

To identify clusters, we use multivariate visualization methods, which imply "optimally" restructuring the rows and columns of the matrix S and puts banks with a similar balance sheet or diversification in loan portfolio closer to each other and different banks far from each other.

We considered another standardized measure for clustering objects based, hierarchical clustering, which does not require predetermining the number of clusters (k) to be computed, and this characteristic makes it the best measure in grouping clusters. One type of this measure is agglomerative clustering in which, each data point is primarily taken as a cluster of its own (leaf). After, the most alike clusters are continuously merged till there is only one single cluster called the root. The final result of hierarchical clustering is a graphical representation in a tree-based form of the objects, commonly known as dendrogram. To visualize the clusters, we used the R package dendextend.

5. RESULTS

In this present paper, we employed the above presented methodology focusing on the balance sheet assets, loan portfolio, and loans adjusted with NPLs ratios as a risk measure. The analysis shows that assets expose several commercial banks to shared risks. On the other hand, the loan portfolio also echoes the message from the assets as some banks present the same similarities as in assets. When we adjust the loan portfolio with risk element, the separation of the hitherto identified clusters of banks becomes even more pronounced. Both empirical methods turn out to be useful matches and reveal how credit risk can concentrate on banks with a similar source of funding and credit exposures to the real economy.

5.1. Banking Sector loans share and concentration

Table 1 below provides an overview of Rwandan banks' exposure to the different economic sectors and suggests that dynamics across sectors have been mixed. Since 2010, mortgage sector has had the highest share, and continues to get bigger as by end 2019. It had 37% of total outstanding bank loans, higher than its last ten years average (32.5%). The construction sector continued to expand with many big projects especially in Kigali City and leads in terms of loan concentration. Other main sectors include trade, manufacturing, transport, and warehousing, personal (consumer) loans, and lastly, hotels and restaurants. Among these, the share of personal loans and loans to the trade sector has declined recently while manufacturing and transports sector share increased. The share of hotels and restaurants has generally remained stable.

Table 1: Banking sector loans share and concentration per sector

Sectors	Share in total loans (average 2010-2019)	Share in total loans in 2019	HHI(average 2010-2019)	HHI in 2019
Mortgage industries	0.325	0.37	0.168	0.176
Trade*	0.175	0.145	0.137	0.152
Manufacturing activities	0.088	0.12	0.204	0.179
Transport & warehousing	0.081	0.112	0.181	0.306
Personal loans	0.126	0.075	0.209	0.18
Restaurant & hotel*	0.082	0.073	0.272	0.253
Water & energy activities	0.022	0.052	0.5	0.47
Service sector	0.034	0.029	0.197	0.139
Agricultural, fisheries & livestock	0.026	0.012	0.284	0.257
OFI & Insurance	0.013	0.011	0.406	0.41
Mining activities	0.001	0.001	0.775	0.778

*Average for trade sector and restaurant and hotels start in 2016 as the two sectors were separated that time

Source: National Bank of Rwanda and Authors' calculation

Regarding concentration, the analysis of HHI per sector helps us to unmask important implications. As discussed in the previous section, for the last ten years, the market for loans had remained unconcentrated, except in the last two years, where the level of concentration moderately increased. The HHI per sector in table 1 shows strong heterogeneity across the sector in terms of loan market concentration, as the latter is exceptionally high for the mining sector, transport and warehousing sector, water and energy sector, and other financial institutions and insurance sector. High HHI in those sector means that the big chunk of loans directed to those sectors is given by a few banks. Generally, the current level of concentration is closer to the ten years average in most sectors except in transport where concentration is increasing while in the services sector the market recently become unconcentrated.

One positive point to mention with some implications to systemic risks is that in sectors where banking system exposure is higher as shown by sector share in total loans (see table 1), the market is relatively less concentrated except for the transport sector and hotels and restaurant sectors. Nevertheless, for hotels and restaurants sector, the situation is improving as the market became moderately concentrated in 2019 from being highly concentrated in previous years. The fact that sector with higher share is financed by many banks may reduce the severity of impact from a sectoral shock on a given bank balance sheet. It reduces the likelihood of having concentrated exposures to an individual or few borrowers thus limit the maximum loss a bank can incur in case of sectoral shock. Details on the evolution of the share of loans per sector in total banking loans and growth of HHI are displayed in the annexes.

5.2. Similarity in Credit Risk Performance

We primarily give the descriptive statistics to show the basic characteristics of the bank portfolios in terms of credit risk materialization (NPL ratios). The NPL stands for Non-Performing Loan ratio, whereas the CV for the coefficient of variation. The results based on the NPL ratio as a proxy for credit risk materialization are provided in Table 2.

Table 2: Nonperforming loans Mean and Coefficient of variation (2016-2019)

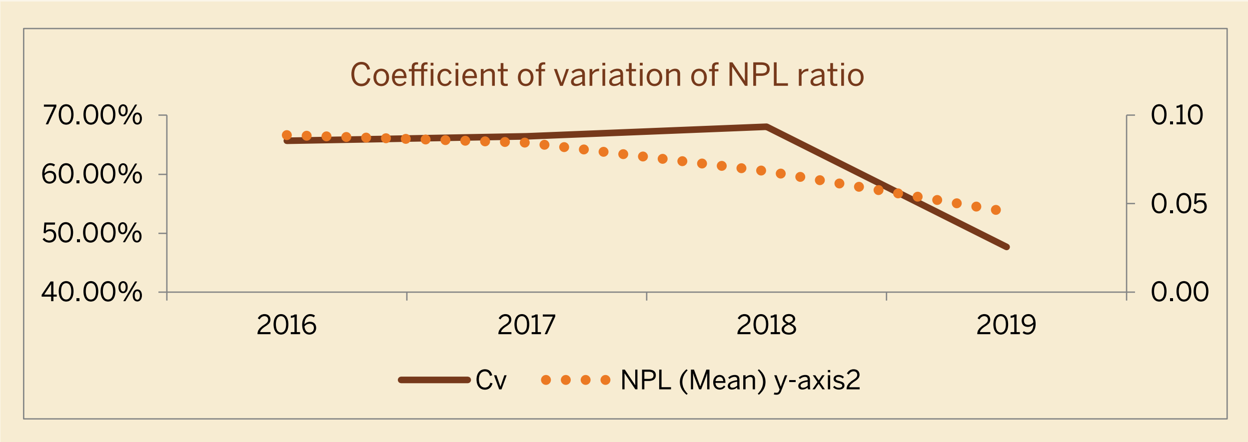
Sector	NPL	CV
Personal loan	0.07	0.98
Agricultural, fisheries& livestock	0.15	1.66
Mining activities	0.02	2.97
Manufacturing activities	0.06	2.19
Water & energy activities	0.01	3
Mortgage industries	0.06	1.28
Trade	0.1	0.88
Restaurant & hotel	0.05	1.86
Transport & warehousing	0.05	1.23
OFI &Insurance	0.03	2.8
Service sector	0.07	1.65

Source: NBR & Authors' calculation

Several sectors present a relatively high level of non-performing loans. In particular, agricultural, fisheries& livestock, trade, service sector as well as personal loans exhibit relatively risky exposures for banks. The risk of indirect contagion via these exposures is even more pronounced given the fact that the coefficient of variation is comparatively small for most of the asset groups.

Conversely, the following sectors, namely water &energy activities, mining activities, and other financial institutions (OFI) & insurance manifest reasonably lower NPL ratios mixed with a high coefficient of variation, thus resulting in relatively low risk. Despite the current low-risk profile of these sectors, the observed low overall dispersion across the banking sector throughout the period under review, as depicted by figure 10, might contribute to a higher level of systemic risk.

Figure 10: Evolution of Non-Performing Loans



Source: NBR and Authors' calculation

5.3. Balance sheet Similarity

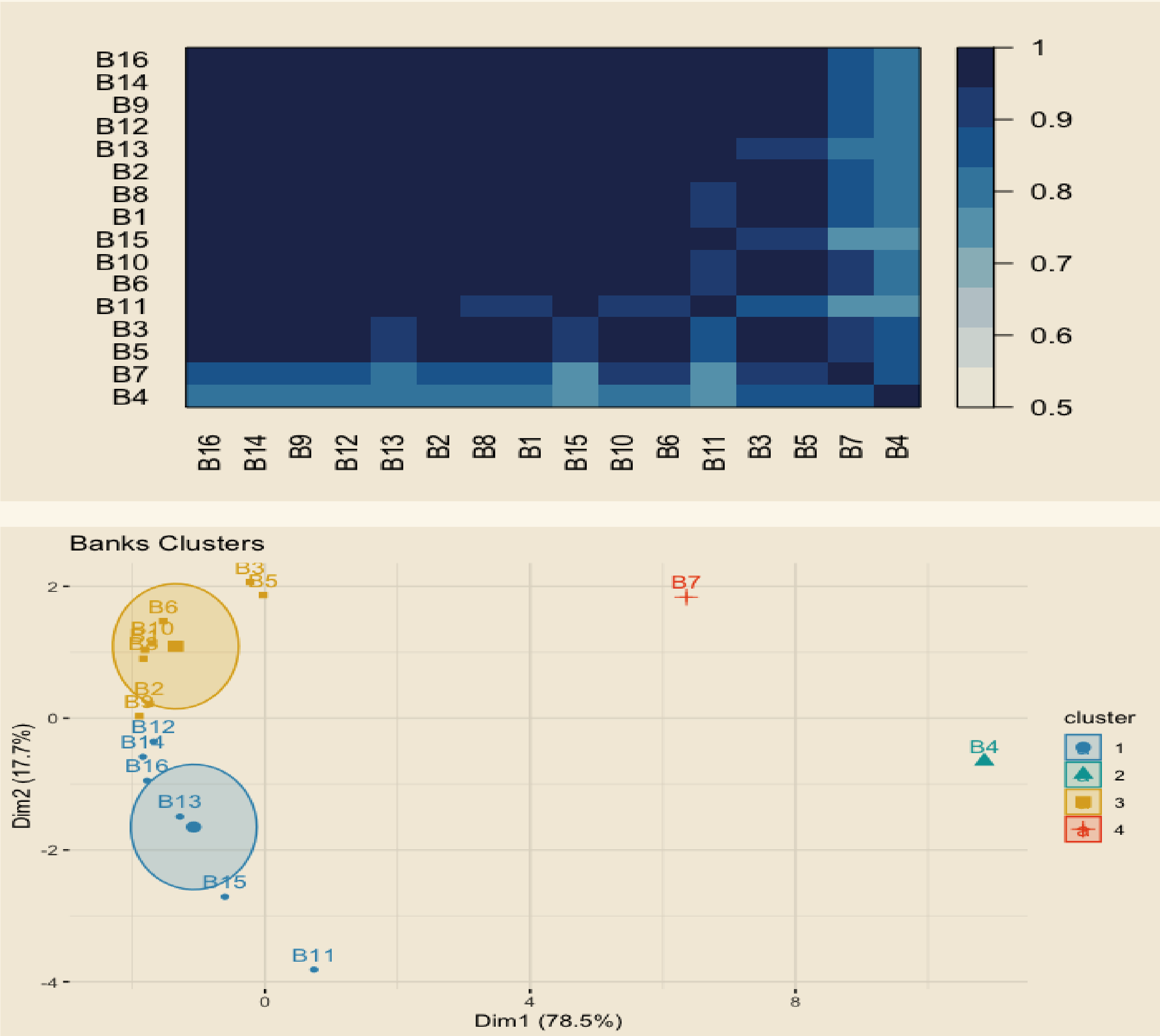
Underlying banks' balance sheet similarity was highly homogenous at 88.35 percent on average in 2019, resulting from observed common assets portfolios for the banks, with the only distinction in the level of asset categories. However, this picture does not fully uncover which banks are most similar and which are less similar. For that reason, Figure 11 displays a similarity matrix for all pairs of banks active as of the end of 2019. Its rows and columns represent individual banks in the same order. Each cell thus stands for the similarity between two banks of the corresponding row and column. The darker blue cell represents the greater similarity between the two banks, while the lighter blue cell symbolizes the lower similarity between two individual banks. The diagonal elements of the similarity matrix show the cosine similarity between a bank and the very same bank, which is equal to 1 by definition.

Using both similarity matrix and cluster analysis, figure 11 shows that two big clusters emerge and the other two small clusters with only one bank in each. Particularly, cluster analysis distinctly describes a yellow-coloured cluster, which contains bank 1, 2, 3, 5, 6, 8, 9 and 10.

The second big cluster (with blue colour) has the subsequent banks, namely banks 12, 14, 16, 13, 15 and 11. The remaining two clusters are considerably small, with bank 7 in the red-coloured cluster and bank 4 in green. Banks are named as numbers as far as confidentiality is concerned.

It is noteworthy to highlight that the largest banks operating in the system mostly belong in the first big cluster, while the smallest banks assembled in the second big cluster.

Figure 11: Banks similarity matrix in the left and clusters in the right as of end 2019



Source: NBR and Authors' calculation

5.4. Similarities between individual banks' loans portfolio

The overall similarity based on loans by sector in the banking system has noticeably loosened from 0.92 in 2016 to relatively stable around 0.63 from 2017 to 2019. Figure 12 in-depth highlights four clusters of similar banks in three consecutive years and quite three clusters in the last period. Also, several clusters invariably remained stable in the period under consideration. These include the following: the cluster of banks 12, 14 and 15; the cluster of banks 3, 5, 6, 7, 11 and a group of banks 1, 9, 10 and 13. It is also important to note that banks 4 and 16 were in the same cluster in the contemporary two years.

Interestingly, a stable cluster of banks 12, 14 and 15 seemed to be the subset of the observed cluster of small banks considering balance sheet size, while the cluster of banks 1, 9 and 10 includes some of the largest banks. This reflects that these groups of banks share the same exposures in underlying loans bestowed in different sectors, namely agriculture, mining, trade, and so forth.

The risk-adjusted similarities based on credit exposures to the real economy, however, portray a different but meaningful picture of banks' similarity. To that end, overall banks' similarity was somewhat unchanged in the first three years at 0.75 and peaked to 0.79 in 2019. As a result, this unmasks the increasing similarity in non-performing loans in particular sectors.

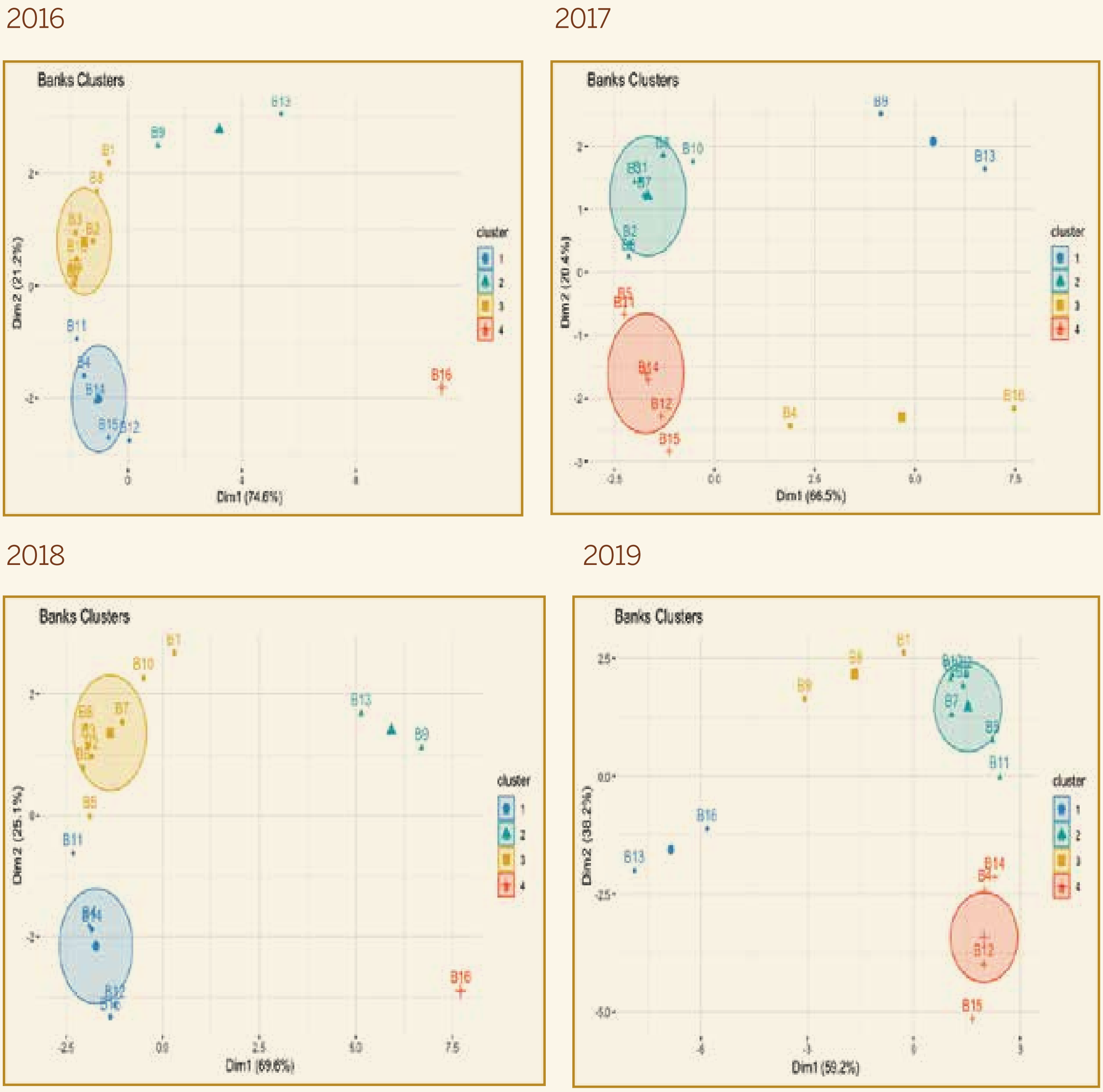
Looking at figure 13 that describes risk-adjusted similarity clusters, the number of clusters stayed stable at four in the period under review, and specifically, two big clusters are present. Similar to non-risk adjusted similarity, the sub-cluster of banks 12, 14 and 15 persisted the homogeneity in recent four years and more interestingly this cluster has in common the lowest financing of the economic activities, except for the trade and mining financing where they are the second last. This group of banks falls in small-sized banks that are less important individually. Even though their total size taken as a group is still relatively small, in event of a common shock to a sector they have lent to, a simultaneous crisis in 3 banks could be seen as a bad sign for financial system stability.

In addition, within clusters, there emerged another clique of five banks that moved together underlying high similarity in 2016-2019, namely banks 2,3,6,7 and 10. This clique has the properties of being the second-lowest lender in the banking system, and they almost finance every activity at an average level compared to other clusters. It is worth noting that banks 1, 8 merged with this cluster in 2016-2018, however, in 2019 moved to a separate cluster of big lenders in every economic sector, except for agriculture, where they are the second largest. The aforementioned group of banks include large and mid-sized banks, thus it deserves keen attention as some of them are already big, and while considered in as a group they are systemically important.

Figure 12: Clusters based on loans by sector

Source: NBR and Authors' calculation

Figure 13: : Risk-Adjusted Similarities between individual Banks

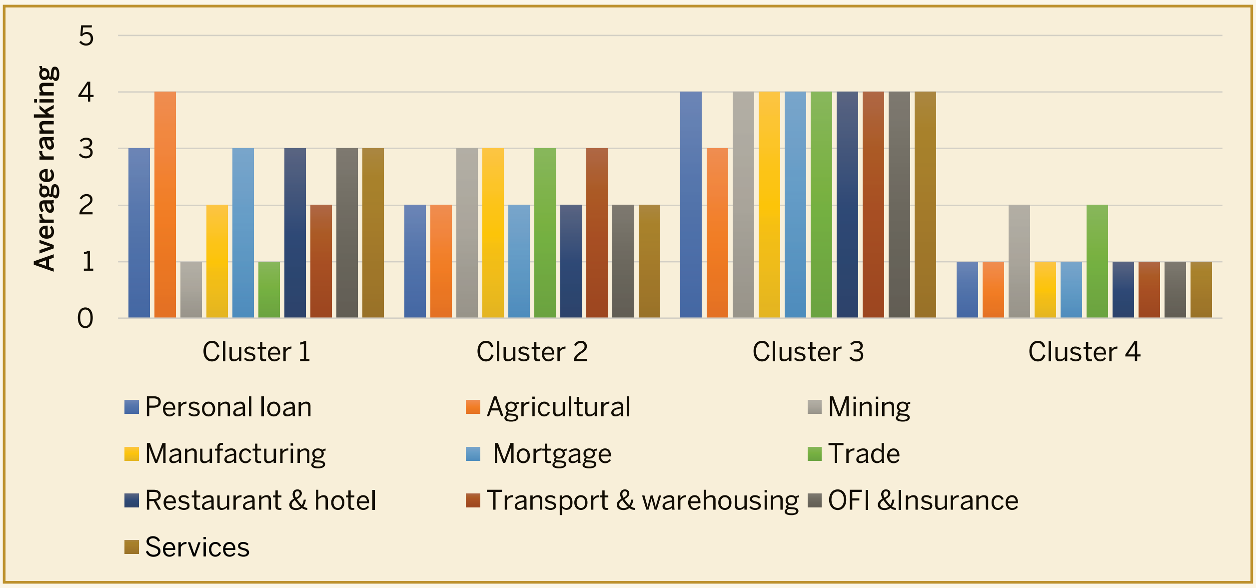


Source: NBR and Authors' calculation

To understand the characteristics of the clusters based on the loans with risk-adjusted for the year 2019, we computed the sectoral average loans and ranked clusters accordingly. As depicted in figure 14, cluster one has the banks that finance the economic activities moderately, except for agriculture, where they are the first lenders. The banks in cluster two are the second-lowest lender in the banking system, and they almost finance every activity at an average level compared to other clusters. In contrast, cluster 3 is made of big lenders in every economic sector, except for agriculture, where they are the second largest.

In cluster 4, we found the lowest financing group of the economic activities, except for the trade and mining financing where they are the second last.

Figure 14: Clusters' features



Source: NBR and Authors' calculation

6. CONCLUSION

In the present study, we assess the systemic risk of banks using portfolio similarity and clustering methods using data from Rwandan banking sector between 2016 and 2019. The main insight is that similarity and cluster analysis techniques help uncover the groups of banks that share the same features in assets and loans portfolio. For as much as banks have common features, they are prone to the same risk exposures, thus the failure of one bank due to one sector could fail similar banks. Therefore, this study provides additional tools to macro-prudential policymakers especially with regards to the assessment of systemic risk.

Results suggest that the overall similarity based on loans by sector in the banking system was noticeably high at 90 percent in 2016 and relatively stable at 63 percent in the following periods. However, adjusting loans with risk-weight based on credit exposures to the real economy, the findings reveal that overall banks' similarity was somewhat unchanged in the first three years at 75 percent and peaked to 79 percent in 2019. Hence, this points out to increasing similarity in non-performing loans in some sectors.

For the year 2019, the banking sector has four clusters. In the first cluster, banks finance the economic activities moderately, except for agriculture, where they are the first lenders. The banks in cluster two are the second-lowest lender in the banking system, and they almost finance every activity at an average level compared to other clusters. In contrast, the third cluster is made of big lenders in every economic sector, except for agriculture, where they are the second largest. In the fourth cluster, we found the lowest financing group of the economic activities, except for the trade and mining financing where they are the second last.

We found a group of five banks including large and mid-sized banks, which have been similar for 3 years and which become even more systemically important as a group, hence, would deserve more scrutiny from macro-prudential policymakers, especially under certain circumstances such as shocks to the common sector they have lent to. Also, there is another outstanding cluster of three small banks which are less important individually but have been very similar in many aspects over the sample period. Even though their importance as a group is relatively small in the Rwandan banking system, this cluster should also be watched closely as its high similarity implies that a shock to assets portfolio of one bank would quickly and simultaneously spread to other in the cluster.

This study provides an additional tool for identifying systemic vulnerabilities from a different perspective. It shows that some medium-sized banks have been consistently similar in terms of the loan portfolio and associated risks in the last four years hence they can be exposed to common risks with impactful consequence as the cluster is sizeable than banks have taken individually. This can be of interest for macro-prudential authorities because, in addition to considering bank size or centrality in interbank transactions or individual bank exposure to a given sector, dynamics in the identified cluster are equally important for systemic stability.

As a policy implication, the National Bank of Rwanda, as a regulatory and supervisory body of commercial banks, should also consider using banks' similarity and clustering tool to carry regular monitoring of systemic risks in addition to other tools such as stress tests and market analysis. This tool can be used, preferably on quarterly basis, to check the individual banks, which can be simultaneously affected by a common shock, and might become systemically important if taken together as a cluster. This can help detect the systemic risk early and guide authority in taking appropriate policy measures such as prudential requirements to limit some banks' excessive exposure to a given sector(s) or to make that exposure more secure, to alleviate risks to financial system stability.

Some potential areas of future research include considering other features of banks' balance sheet especially on liabilities side such as banks funding structure or banks liquidity, which will add more insights regarding banks similarity and implications on systemic stability. Secondly, identifying the drivers behind bank similarity could also help to understand the phenomenon.

REFERENCES

Aldasoro, I. & Alves, I., 2016. Multiplex interbank network and systemic importance: an application to European data, Frankfurt: ECB Working Paper Series No. 1962.

Andrea A. Jobst and Dale F. Gray, 2016. Systemic Contingent Claims Analysis - Estimating Market - Implied Systemic Risk. International Monetary Fund, 13(54), p. 78.

Balazs C., 2018. Bank Network Analysis in the ECCU. IMF, 18(162), p. 32.
Bar-Joseph, Z., Gifford, D. K. & Jaakkola, T. S., 2001. Fast optimal leaf ordering for hierarchical clustering. Bioinformatics, 17(1), pp. S22-S29.

Boitan, N. D. a. I. A., 2009. A Cluster Analysis Approach for Banks' Risk Profile: The Romanian Evidence. European Research Studies, XII(1), p. 10.

Brechler, J., Hausenblas, V., Komárková, Z. & Plašil, M., 2014. Similarity and Clustering of Banks: Application to the Credit Exposures of the Czech Banking Sector, s.l.: Czech National Bank.

Brechler, J., Hausenblas, V., Komárková, Z. & Plašil, M., 2014. Similarity and Clustering of Banks: Application to the Credit Exposures of the Czech Banking Sector, s.l.: Czech National Bank.

Caccioli, F., Shrestha, M., Moore, C. & Farmer, J. D., 2014. Stability analysis of financial contagion due to overlapping portfolios. Journal of Banking and Finance, Volume 46, pp. 233-245.

Cai, J., Eidam, F. & Saunders, A. S. S., 2018. Syndication, Interconnectedness, and Systemic Risk. Journal of Financial Stability, 34(C), pp. 105-120.

Chretien, E. et al., 2020. Exposure through common portfolio and contagion via bilateral cross-holding among funds, banks and insurance companies, Paris: Haut Conseil de Stabilité Financière.

Chuanmin, M., Runjie, X. & Qingtong, L., 2019. Systemic Risk Clustering of China Internet Financial Based on t-SNE Machine Learning Algorithm, s.l.: https://arxiv.org/abs/1909.03808v1.

Csonto, B., Guerson, A., Myrvoda, A. & Sewordor, E., 2018. Bank Network Analysis in the ECCU, Washington DC: IMF Working Paper No. 18/162.

Dardac, N. & Boitan, I. A., 2009. A Cluster Analysis Approach for Banks' Risk Profile: The Romanian Evidence. European Research Studies Journal, XII(1), pp. 109-118.

Drehmann, M. & Tarashev, N., 2011. Measuring the systemic importance of interconnected banks, Basel: BIS Working Paper No 342.

ECB, 2010. Finacial Stability Review, Frankfurt: European Central Bank.

Espinosa-Vega, M. & Sole, J., 2010. Cross-border financial surveillance: A network, Washington DC: IMF Working Paper No. 10/105.

Gai, P., Haldane, A. & Kapadia, S., 2011. Complexity, Concentration and Contagion. Journal of Monetary Economics, 58(5), pp. 453-470.

IMF, BIS and FSB, 2009. Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations, Washington DC: International Monetary Fund.

Josef Brechler, V. H. Z. K. a. M. P., 2014. Similarity and Clustering of Banks: Application to the Credit Exposures of the Czech Banking Sector. Policy and Research notes, Issue 4, p. 30.

Krause, A. & Giansante, S., 2012. Interbank lending and the spread of bank failures: A network model of systemic risk. Journal of Economic Behavior & Organization, Volume 83, p. 583– 608.

Leon, C., 2017. Banks in Colombia: how homogeneous are they?, Bogota: Banco de la Republica Colombia.

Leon, C., Machado, C. & Murcia, A., 2015. Assessing systemic importance wtha fuzzy logic inference system. Intelligent system in accounting, finance and management, 23((1-2)), pp. 121-153.

Leon, C., Machado, C. & Sarmiento, M., 2018. Identifying central bank liquidity super-spreaders in interbank fundsnetworks. Journal of Financial Stability, 35(C), pp. 75-92.

Lux, T., 2016. A model of the topology of the bank-firm credit network and its role as channel of contagion, s.l.: Kiel Working Paper, No. 1950.

Mi Chuanmin, X. R. L. Q., 2019. Systemic Risk Clustering of China Internet Financial Based on t-SNE Machine Learning Algorithm. Nanjing University of Aeronautics and Astronautics, p. 10.

Minoiu, C. & Reyes, J. A., 2011. A network analysis of global banking: 1978-2009, Washington DC: IMF Working Paper No. 11/74 .

NBR, 2019. Annual financial stability report July 2019-June 2019, Kigali: National Bank of Rwanda.

Nyalihama, C. & Kamanzi, C., 2019. Financial sector developemnt and impact of economic growth in Rwanda. BNR Economic Review, Volume 15, pp. 1-56.

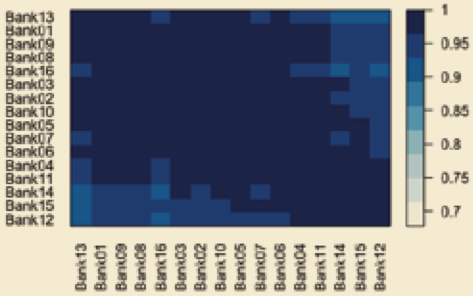
Sheldon, G. & Maurer, M., 1998. Interbank Lending and Systemic Risk: An Empirical Analysis for Switzerland. Swiss Journal of Economics and Statistics, 134(IV), pp. 685-704.

Tabak, B. M., Takami, M., Rocha, J. M. C. & Cajueiro, D. O., 2011. Directed Clustering Coefficient as a Measure of Systemic Risk in Complex Banking Networks, Brasilia: Banco Central do Brasil Working Paper Series 249.

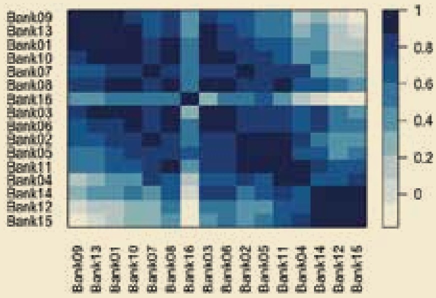
APPENDICES

1. Loans based similarities Between Individual Banks

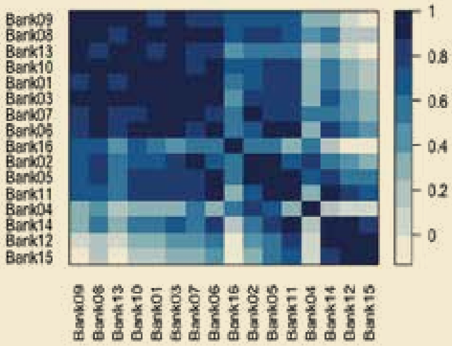
2016



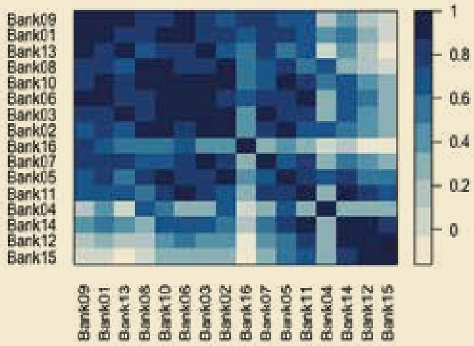
2017



2018



2019



2. Risk- Adjusted Similarities between Individual Banks

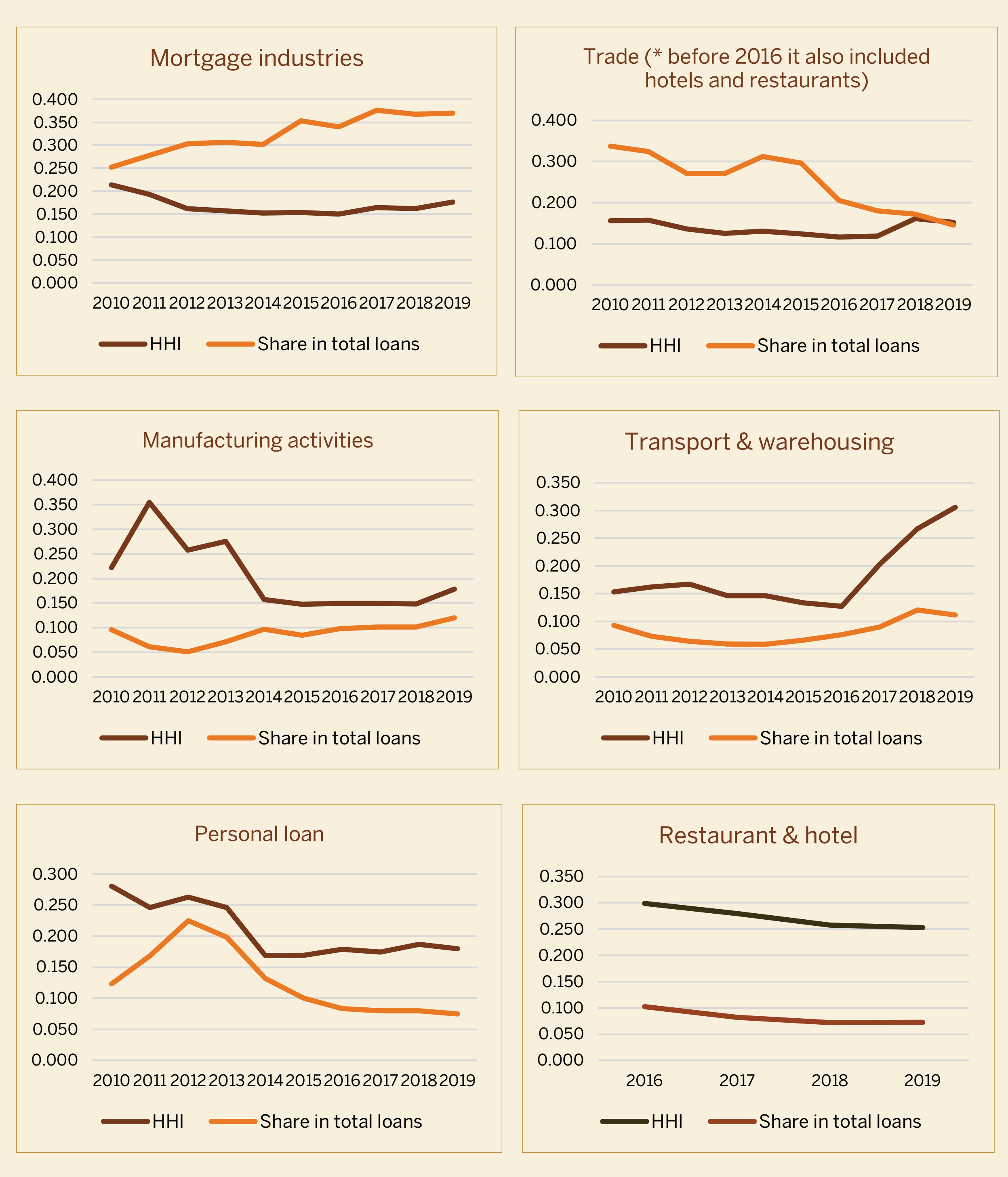
2016

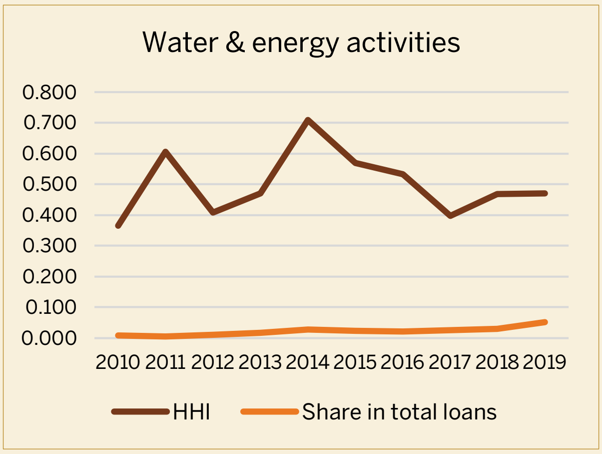
2018

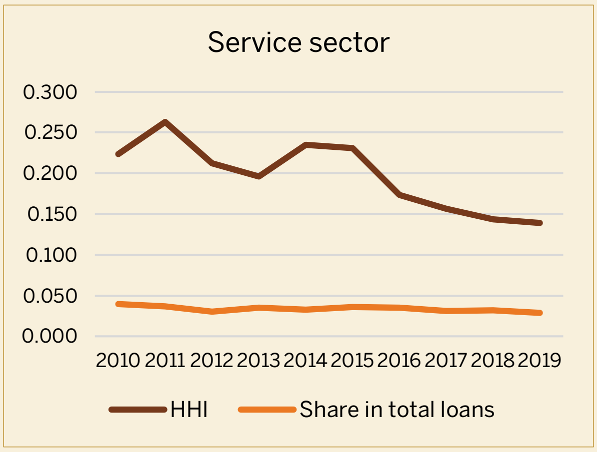
2017

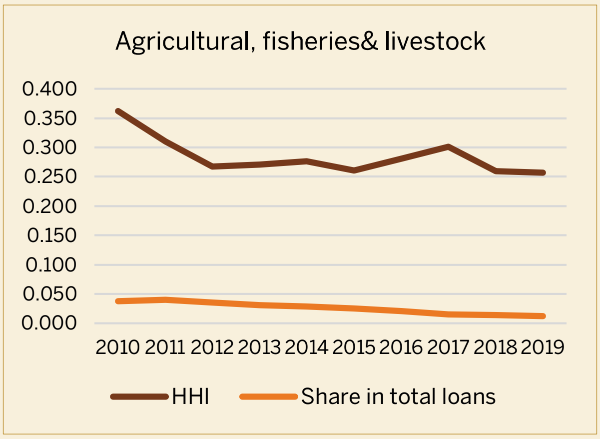
2019

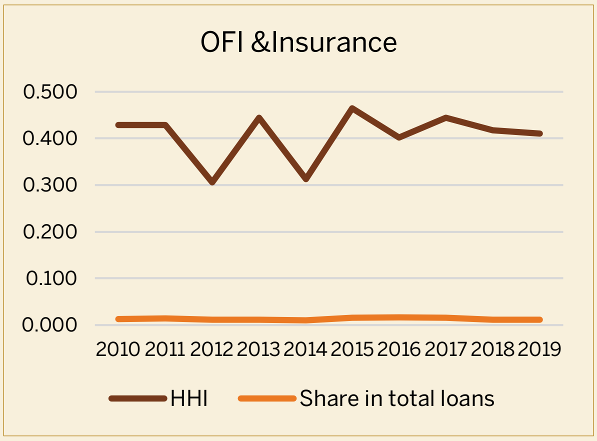
3. Evolution of loans share per sector and HHI per sector

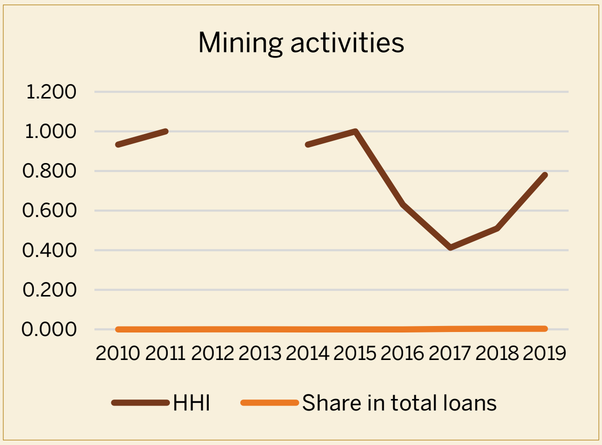



Water & energy activities
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
HHI
Share in total loans


Service sector
0.300
0.250
0.200
0.150
0.100
0.050
0.000
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
HHI
Share in total loans


Agricultural, fisheries& livestock
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
HHI
Share in total loans


OFI &Insurance
0.500
0.400
0.300
0.200
0.100
0.000
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
HHI
Share in total loans


Mining activities
1.200
1.000
0.800
0.600
0.400
0.200
0.000
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
HHI
Share in total loans

FACTORS EXPLAINING LOAN DELINQUENCY AMONG MANUFACTURING FIRMS IN RWANDA: FINANCIAL DISTRESS OR MORAL HAZARD?

Patrick Hitayezu[4]
Sonia U. Kayirere[5]
Alfred Hakuzwimana[6]

[4]Director, Research Department, National Bank of Rwanda
[5]Young Economist, Research Department, National Bank of Rwanda
[6]Economist, Financial Stability Monitoring and Policy Department, National Bank of Rwanda

ABSTRACT

Over the last decade, the Rwandan economy has enjoyed an increasingly sound and stable financial system, despite higher and growing stocks of non-performing loans (NPLs). The manufacturing sector has been among the main contributors to the deterioration of bank asset quality, with the stock of non-performing loans (NPLs) growing from 1 percent in June 2015 to 13 percent in June 2018. Despite their appeal for microprudential policy in Rwanda, micro-level assessments of the determinants of credit delinquency among manufacturing firms remain scanty.

Against this backdrop, this study investigates firm-level determinants of loan delinquency among manufacturing firms in Rwanda using a unique dataset compiled in 2018 from a survey on access to finance conducted among 122 manufacturing firms in Rwanda. Various financial ratios are used to examine whether the surge in NPL was a result of financial distress and/or a moral hazard effect. The results of a logit model show that none of the financial ratios for moral hazard are significant predictors of credit delinquency. In line with the theoretical expectation of the financial distress hypothesis, the results suggest that the odds of credit delinquency significantly decrease with profitability and liquidity ratios. Among firm-specific characteristics, the results indicated that credit default risk increases with firm age. These findings indicate a link between credit default and financial distress, which is in line with the findings of previous studies that suggested the importance of macro-economic performance for loan repayment performance. The study concludes by outlining major implications in terms of banking regulation and financial stability policy in Rwanda.

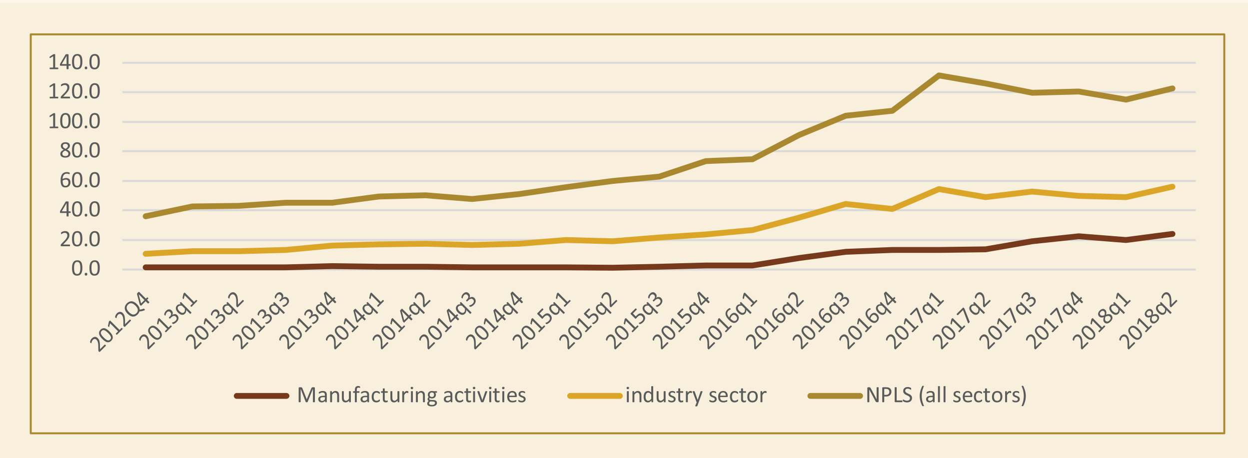
Keywords: credit default risk, moral hazard, manufacturing sector, Rwanda
JEL classification: G33, D82

1. INTRODUCTION

Over the last decade, Rwanda has enjoyed an increasingly sound and stable financial system. With the implementation of the global banking standards (Basel II/III) of the post-2008 global financial crisis (GFC), the financial sector recorded significant improvements in the adequacy of capital buffers, loan coverage, and liquidity, which generally stood above the prudential requirements. Consequently, as the National Bank of Rwanda (NBR, 2020) reports, financial intermediation has remained the core role of the banking sector. Despite these positive developments, higher and growing stocks of non-performing loans (NPLs) remains among the most challenging aspect of bank asset quality. According to various financial stability reports, the ratio of NPL in total loans reached a peak of 8.2 percent in mid-2017 (NBR, 2017) before consistently decreasing as a result of write-offs of overdue NPLs (NBR, 2020).

The deterioration of bank asset quality has also complicated macroprudential policy strategies aiming at reducing the systemic risk related to higher exposure of bank loan books to the real estate industry. According to the latest financial stability report, the gradual diversification of banks' loan books to other sectors such as manufacturing was viewed as a way of mitigating the unhealthy exposure to the mortgage sector (NBR, 2020). Incidentally, the manufacturing sector has been among the main contributors to the deterioration of bank asset quality over the last decade. The NPLs ratio in the manufacturing sector grew rapidly from 1 percent in June 2015, to 6 percent in June 2016 and 8 percent in 2017. The ratio of NPL in this loan portfolio reached a peak of 13 percent in 2018. As the figure below shows, the stock of NPLs from the manufacturing sector was around FRW24 billion in the second quarter of 2018, from under 8 billion in the corresponding period of 2016.

Figure 1: Non-performing loan volumes (stock)



Source: Author's compilation from NBR data.

Beyond the implications for financial stability, the deterioration of credit score among manufacturing firms has implications for broader development policy in Rwanda. The reduction in the sector's creditworthiness could deter the country's aspiration towards a middle-class economy by 2050. As a recent assessment of the future drivers of economic growth by the World Bank (2020) notes, the sector remains fragile to spur the necessary growth and employment to achieve the country's ambitious vision 2050. The report argues that the sector requires an aggressive mobilization of massive resources to boost productivity.

Addressing this policy challenge, therefore, requires a deep understanding of the drivers of credit default among manufacturing firms. Empirical research into this field, however, remains scanty. To the best of the authors' knowledge, only a recent study by Karuhanga et al. (2018) unveiled macro-level determinants of credit default risk in Rwanda, showing that NPLs ratios in commercial banks of Rwanda increase with inflation, exchange rate depreciation, and real interest rate, and decreases with bank size and GDP growth rate.

The negative effect of economic growth on NPLs revealed by Karuhanga et al. (2018), in particular, is in line with the predictions of the financial distress hypothesis that link credit default probability to a firm's financial distress (Altman, 1968; Ciampi, et al., 2021). However, recent advances in credit default prediction modeling have shown that financial distress is not the sole cause of credit default among borrowers. Based on the notion that some borrowers may find it economically more attractive not to default on their liabilities (a decision referred to as strategic default), there is growing empirical evidence showing the predictive power of moral hazard on credit delinquency of households (Guiso, et al., 2013) and enterprises (Asimakopoulos, et al., 2016; Giroud, et al., 2012; Hyytinen & Väänänen , 2006; Castillo, et al., 2017). A number of studies have shown that loan default rates increase with the extent of unobservable moral hazard (Asimakopoulos, et al., 2016) as well as its observable indicators (Castillo, et al., 2017). The distinction between financially-distressed defaulters and strategic defaulters has direct implications for banking regulation and financial stability policies. The later form of credit default usually necessitates extra credit risk management systems to close the information gap between lenders and borrowers. Micro-level studies are therefore necessary to shed light on the prevalence of moral hazard among defaulters to inform on the appropriate course of policy action.

This paper, therefore, seeks to fill the knowledge gap by assessing firm-level determinants of credit default probability among Rwandan manufacturers. We contribute to the existing literature by specifically investigating the effects of unobserved factors and observable indicators of moral hazard on a firm's credit default risk. We use a unique dataset compiled from a cross-sectional survey on the access to finance conducted among 122 manufacturing firms in Rwanda in 2018, i.e. the year when NPLs from the manufacturing sector peaked. The dataset contains the details of balance sheet and income statements of the surveyed manufacturing firms in 2018.

The performance of the manufacturing sector of Rwanda on the credit market offers a unique field for observing the moral hazard effect. A recent study by Hitayezu et al. (2019) showed that the institutional environment under which the credit market operates is characterized by inadequate information sharing. They show that the information asymmetry between banks and manufacturers leading to substantial credit rationing by banks. It is therefore desirable to investigate the extent to which information asymmetry determines the performance of credit market ex-post in the form of moral hazard.

The rest of the paper is organized into four sections. The subsequent section briefly reviews the literature on microeconomic determinants of credit delinquency and briefly presents existing empirical evidence. The third section outlines the methods and material used. Section 4 presents an analysis of survey data with design weights, provides the model result, and discusses the main findings. The last section summarizes the main findings and outlines key policy implications.

2. LITERATURE REVIEW

Credit delinquency has been intensively studied within the framework of information economics. Under this strand of research, the principal-agent theory posits that information asymmetry between the principal (i.e. the lender) and the agent (i.e. the borrower) on the credit market leads to adverse selection and moral hazard problems (Stiglitz & Weiss, 1981). The moral hazard problem arises from the possibility that the borrower may take actions that may not be observable to the lender ex-ante.

Despite the appeal of the moral hazard hypothesis, empirical investigation into the moral hazard effect has been a challenging endeavor. Facing the challenges related to the measurement of moral hazard, the early credit default prediction modeling literature was mainly confined to the notion that financial distress was the principal predictor of credit default risk (Ciampi, et al., 2021). Most of these studies reported that a firm is more likely to default if it is highly indebted, less profitable, or less liquid (Altman, 1968; Ciampi, et al., 2021).

Recently, however, especially with the increasing availability of richer datasets, the specific effect exerted by the moral hazard has come to the fore in the empirical literature. Indeed, authors have argued that the effects of some of the financial ratios on credit default risk as in Altman(1968) can indeed vary depending on whether the default is a consequence of financial distress or a strategic decision prompted by its economic attractiveness. The effects of two financial ratios in particular, viz. indebtedness and profitability, have been extensively investigated. Economists argued that highly indebted firms pay higher shares of their earnings to lenders, which could reduce the incentives to make efforts to increase the success of the project (or rather increase the incentives to venture into riskier projects) (Fidrmuc & Hainz, 2010). With regard to profitability, economists have asserted that though higher margin and return ratios indicate a firm's ability to pay back debts, firms with higher shares of retained earnings have the ability to run their businesses without external funding, leading to higher propensities to default on their loans (Asimakopoulos, et al., 2016).

The literature suggests that firm-specific factors could mediate the effect of moral hazard on loan repayment behavior. Some economists have argued that moral hazard could decrease with the degree of borrower's liability. They asserted that debtors with full liability internalize the positive and negative outcomes of their investment decisions in their payoffs (which lead to a better repayment performance), while debtors with limited liability would prefer to repay only in the case of successful payoffs (especially when they have insufficient assets to be liquidated) (Fidrmuc & Hainz, 2010). Firm size and age may also mediate the effect of moral hazard on strategic default risk. In terms of firm age, Diamond (1989)'s model indicates that borrowers rely on building positive reputations for repayment, suggesting that older firms would have little incentive to default on their loans. The age effect, however, is thought to be non-linear.

Smaller and younger firms with higher switching costs may be inclined to avoid actions that could impair their relationship with the lender (a phenomenon known as hold-up effect) which effectively mitigates moral hazard, while larger and well-established firms would be reluctant to engage in loan delinquency to avoid tarnishing their reputation (which they use to secure low-cost funding) (Asimakopoulos, et al., 2016).

With regard to collateral value, the literature suggests two contrasting effects of collateral on credit default risk. From an ex-ante perspective, a negative relationship between collateral and default risk emanates from a borrower selection effect, as good borrowers choose to pledge collateral to signal their underlying quality and get lower risk premiums. In an ex-post fashion, however, a positive relationship between collateral pledging and default rates stems from three moral hazard effects. These include the lender selection effect (banks require low-quality borrowers/projects to pledge collateral to reduce moral hazard and other ex-post frictions and protect themselves from loan risks), the risk-shifting effect (borrowers' motivation to shift into a safer investment portfolio to avoid asset loss) and the loss mitigation effect (collateral pledges reduce losses for the lender in the event of borrower default) (Le & Nguyen, 2019).

The hypotheses above have been explored in the empirical literature. Asimakopoulos et al. (2016), for example, found that profitability and collateral can be used to distinguish strategic defaulters from financially distressed defaulters in Greece. Their logit regressions revealed a positive relationship of strategic default with outstanding debt and a negative relationship with collateral value. They further report an inverse U-shaped relationship between strategic default risk and firm size and age.

Although moral hazard (as described above) is not directly observable by the lender, its outcomes can be recognizable. With this in mind, some recent studies have attempted to provide more observable indicators of moral hazard that could help lenders monitoring their borrowers' behaviors. Castillo et al. (2017) argue that the problem of moral hazard that may arise from the borrower side can be illustrated by looking at the financial performance of SMEs with collateralized loans. For example, alternating use or investing financial resources in different projects arranged with the lender is argued to be an indicator of moral hazard (Bebczuk & Bebczuk, 2003). Also, the problem of asset substitution may arise when borrowers willingly deceive lenders after a financial transaction has occurred, i.e. after the loan has been issued, by replacing higher-quality assets with one with relatively lower quality (and price) or rather with a riskier one. The lender in this case may not know if the borrower used the funds for the purpose they were intended for. Simply put, the asset substitution problem highlights the conflicts between shareholders or borrowers and creditors, whereby borrowers try to shift the risk to creditors (Castillo, et al., 2017).

Using different financial ratios as indicators of characteristics of moral hazard (asset substitution, low efforts, underinvestment, and different use of invested capital), Castillo et al. (2017)'s logit models showed that SMEs in Colombia that had higher scores of low effort and underinvestment had a significantly higher propensity to default on their credit. The crux of the literature briefly reviewed above is that financial distress indicators alone cannot explain credit default risk among SMEs. Despite their difficult measurement, unobservable indicators and observable characteristics of moral hazard are worth-considering in models of credit repayment performance.

3. METHODOLOGY

This section discusses the empirical model adopted in the study, source of data, model specification, and variables used to analyze firm-based factors influencing credit constraint among manufacturers.

3.1. Empirical model

The general model employed in the empirical framework utilizing different specifications and estimation methods is represented as follows:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (1)$$

Where Y_i represents a measures of loan delinquency; X_i the vector of variables representing different financial ratios and firm characteristics; α, β, parameters to be estimated; and ε_i the random error. For estimation purposes, we apply Logit model. Previous studies have shown that Logit models yield higher predictive power for measuring the effect of financial distress on credit default risk than probit and tobit models (Altman & Sabato, 2007). Logit models have been extensively applied in the moral hazard and financial distress literature. The model was also applied in recent studies investigating the effects of both financial distress and moral hazard on credit default risk such as Asimakopoulos et al. (2016) and Castillo et al. (2017). In our probability model, a binary outcome variable Y_i will represent the presence/absence of irregular repayment in 2018.

3.2. Data sources

This study uses data from a survey on access to finance of industrial firms that was conducted by the National Bank of Rwanda. The data were collected from a sample of 122 manufacturing firms. The sample selection followed a 2-stage cluster sampling approach, using a long list of registered manufacturing firms provided by the National Institute of Statistics of Rwanda (NISR). In the first stage, 12 divisions of activities were purposively selected from the long list of registered manufacturing firms, based on the International Standard Industrial Classification (ISIC). In the second stage, firms were randomly selected using a simple random sampling method.

The sample covered the city of Kigali and all the provinces. Outside Kigali, interviewed firms were from various districts including Rubavu, Rwamagana, Nyagatare, Nyaruguru, Nyamasheke, and Rulindo.

The survey used a structured interview questionnaire. The survey questionnaire was structured into three main sections. The first section focused on firm-specific details (including registration, ownership, management, and performance). The second section included questions on firm financing, including relevant sources of internal and external financing, perceptions about the banking sector and capital market and experience with credit market. To avoid issues related to recall bias, only specific details of credit history in 2018 were captured. The history included loan application and outcome, loan size and use, and repayment performance. The last section of the survey questionnaire captured information from two financial statements of firms namely, balance sheet and income statement. The questionnaire was coded in SurveyCTO which was the platform used to collect data.

The questionnaire was administered using face-to-face interviews. A team of 12 enumerators composed of Young Economists from the National Bank of Rwanda was trained beforehand. After pilot-testing the instrument, face-to-face interviews took place from May to June 2019, using tablets. Interviews took place within the premises of firms. Respondents were primarily senior managers of firms with good knowledge of firm financing. A consent form was included in the questionnaire, and respondents were asked at the beginning of the face-to-face survey if they agreed to be a part of the study. On average, an interview took 1 hour and 21 minutes.

3.3. Model specification

3.3.1. Dependent variable (IRREGULAR)

In line with the Basel Committee on Banking Supervision's guidance that defines non-performing loan based on the criteria of delinquency status, this study classified firm's repayment performance on current and past loans into two categories: the regular repayment category and the irregular repayment category. The latter included borrowers that had made zero payments of interest or principal within 90 days or their payments were 90 days past due (overdue). This categorization was used to compute a binary outcome variable named IRREGULAR. The dummy variable took the value of 1 for a firm that had been in the NPL category (for the current or past loans) and 0 otherwise.

3.3.2. Independent variables

To investigate the potential effect of financial distress on loan repayment performance (as in (Altman, 1968)), the model included three traditional financial ratios, namely liquidity, profitability, and leverage ratios.

Table 1: Traditional financial ratios used in the empirical model

Variable name	Measurement	Theoretical expectation
LIQUIDITY	Cash over total assets	Expected effect on financially distressed default (-): A less liquid firm is expected to experience significant cash-flow problems that hinder its ability to abide by its loan repayment schedule.
PROFITABILITY	Earnings before Interest, Tax, Depreciation and Amortization (EBITDA) over total asset	Expected effect on financially distressed default (-): management efficiency in utilizing the firm's assets to generate profits was expected to decrease with the probability of default. Expected effect on strategic default (+): a higher share of retained earning provides the ability to run a business without external funding, leading to a higher propensity for strategic default.
LEVERAGE	Debt-to-equity ratio	Expected effect on strategic default (+): Highly indebted firms pay higher shares of their earnings to lenders, which could reduce the incentives to make efforts to increase the success of the project (or rather increase the incentives to venture into riskier projects) (Fidrmuc & Hainz, 2010) or make the option of strategic default more profitable (Asimakopoulos, et al., 2016).

As indicated in the last column of Table 1, the coefficients of most of the traditional financial ratios included in the model could indicate both financial distress and moral hazard effects. Therefore, it is desirable to include financial ratios that specifically indicate potential moral hazard effect. Following (Castillo, et al., 2017), four financial ratios are included in the model as observable indicators of four aspects of moral hazard, namely asset substitution, low effort, underinvestment, and different use of invested capital.

Table 2: Financial ratios for moral hazard used in the binary probit model

Variable name	Measurement	Theoretical expectation
SUBSTITUTION	$\frac{\text{Short-term investments + Long-terminvestments}}{\text{Total assets}}$	**Asset substitution effect (-):** Higher values of SUBSTITUTION indicate an increase in investments in riskier assets made by a firm to secure higher returns in order to meet its obligations.
EFFORT	$\frac{\text{Operating costs and expenses}}{\text{Total assets}}$	**Low effort effect (-):** Lower values of EFFORT indicate that management is decreasing efforts to keep the business afloat.
UNDERINVEST	$\frac{\text{Non-operating expenses}}{\text{Total assets}}$	**Underinvestment effect (+):** higher UNDERINVEST ratio indicates a decrease in investments that are the raison d'être of a company.
DIVERSION	$\frac{\text{Capital expenditure}}{\text{Total assets}}$	**Effect of different use of invested capital (-):** Lower importance of DIVERSIFICATION may indicate that capital is being used differently.

Source: Adapted from (Castillo, et al., 2017)

Three firm-specific characteristics that could mediate the effect of moral hazard were controlled for in the model. These include ownership type, firm size, and land and building value (an indicator of collateral value), as described in Table 3. Given the nature of the size effect outlined in the literature review section, a quadratic form of firm size is included in the model to control for the plausible effect of different age brackets.

Table 3: Selected firm-specific characteristics

Variable name	Measurement	Theoretical expectation
SINGLEOWNER	(Dummy variable) 1= the firm is owned by one investor only, and 0= owned by a partnership of investors (Dummy).	Expected effect on strategic default (-): Debtors with full liability internalize the positive and negative outcomes of their investment decisions in their payoffs (which lead to a better repayment performance), while debtors with limited liability would prefer to repay only in the case of successful payoffs (especially when they have insufficient assets to be liquidated)
FIRMAGE	(Continuous variable) Number of years in the business.	Expected effect on strategic default (±): smaller firms facing higher switching costs may be inclined to avoid actions that could impair their relationship with the lender, while larger and well-established firms would be reluctant to engage in loan delinquency to avoid tarnishing their reputation (which they use to secure low-cost funding) (Diamond, 1989; Asimakopoulos, et al., 2016)
LANDBUILD	(Continous variable) total value of own lands and buildings owned by the firm (in FRW)	Expected effect on strategic default (-): firms with higher value immovable asset selection effect (banks require low-quality borrowers/projects to pledge collateral to reduce moral hazard and other ex-post frictions and protect themselves from loan risks), the risk-shifting effect (borrowers' motivation to shift into a safer investment portfolio to avoid asset loss) and the loss mitigation effect (collateral pledges reduce losses for the lender in the event of borrower default)

4. RESULTS AND DISCUSSION

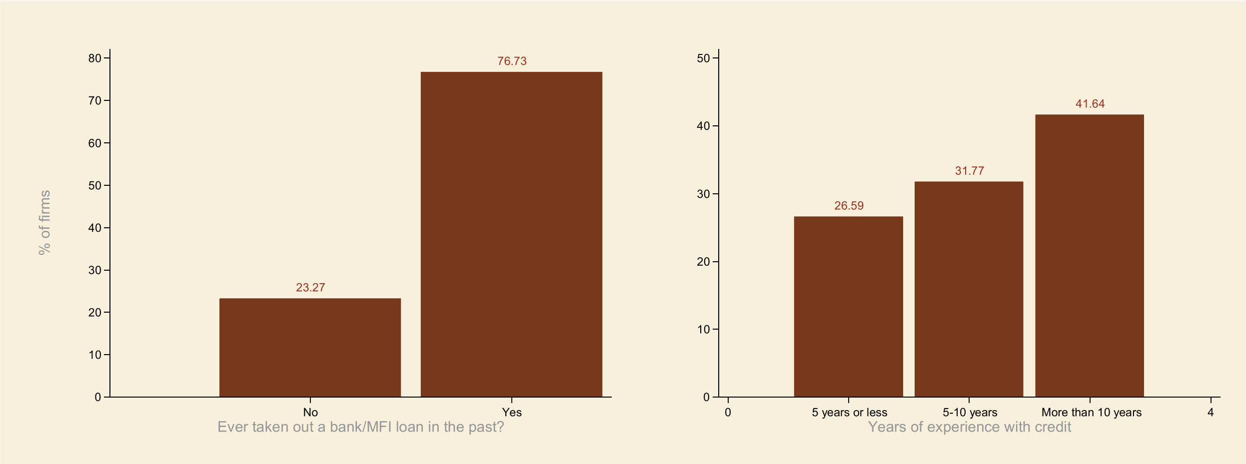
4.1. Analysis of survey data with design weights

The analysis of survey data presented in this section includes design (sampling) weights. This technique permits accounting for unequal selection probabilities. The distributions are presented using histograms.

4.1.1. Firm's experience with bank loan

Bank loan turned out to be an important source of external financing for the majority of manufacturing businesses in Rwanda. The analysis of the survey data provided in Figure 2 suggests that, by 2018, about 76.7 percent of manufacturing firms had taken out a bank loan in the past. Among those with credit history, around 73 percent of firms had five or more years of experience with bank loans.

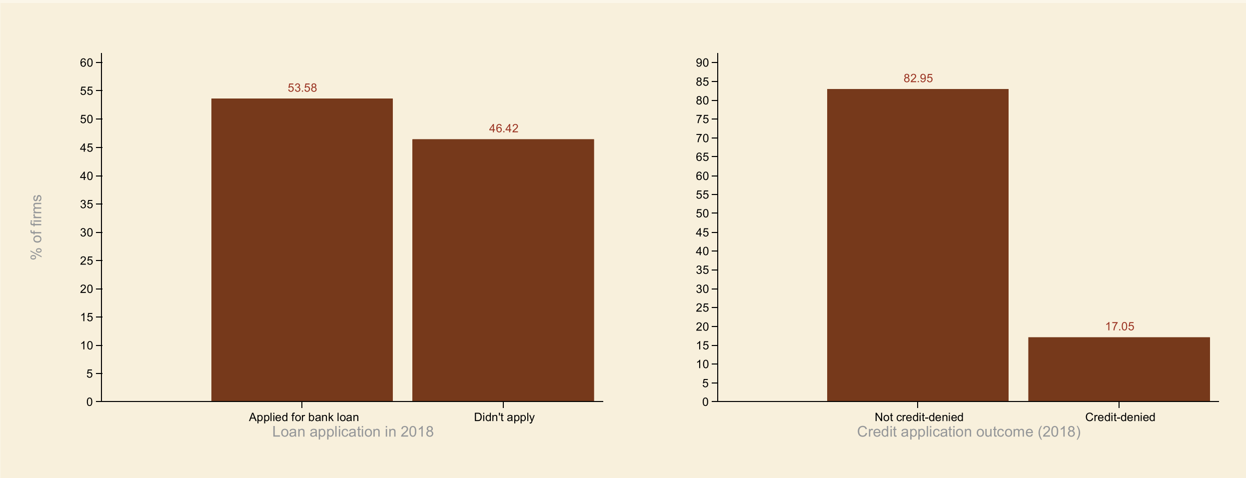
Figure 2: Experience with bank loans among manufacturing firms in Rwanda (2018)



Source: Authors' analysis using survey data

In 2018, the majority of manufacturing firms applied for bank credit, and rejection was not a common experience. Figure 3 suggests that loan application rate stood at 53 percent in 2018. More importantly, more than 80 percent of loan applications were successful.

Figure 3: Loan application and rationing among manufacturing firms in Rwanda (2018)

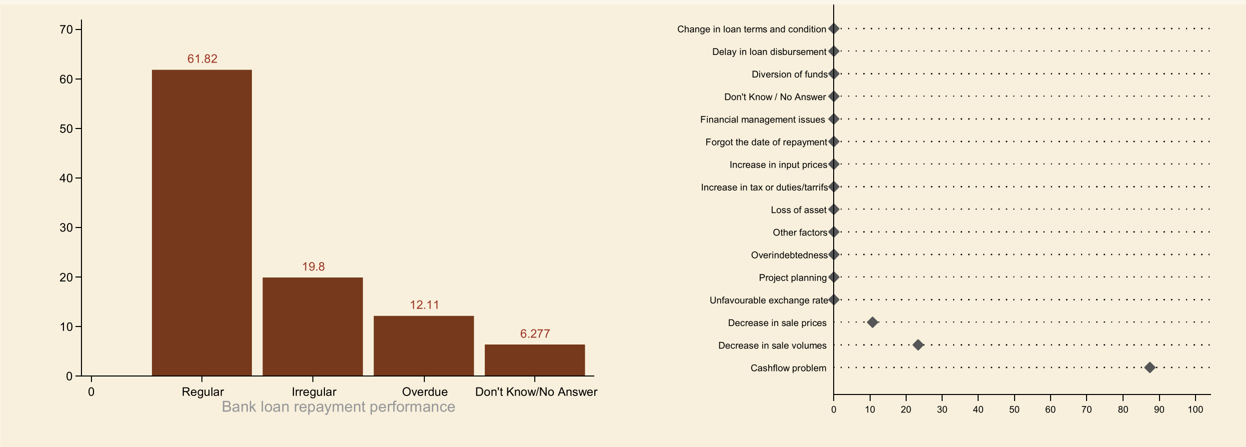


Source: Authors' analysis using survey data

4.1.2. Credit repayment history

Analysis of self-reported loan performance among manufacturing firms in Rwanda suggests substantial delinquency among debtors. Figure 4 shows that about 32 percent of borrowers had histories of irregular or overdue repayments. Analysis of self-reported challenges among firms with histories of delinquent loan account suggests that cash-flow was the most perceived challenge (86 percent), followed by decreased revenues (22 percent) either due to price changes and/or decreased sales. Challenges related to decreasing market prices were perceived by only 10 percent of credit-delinquents.

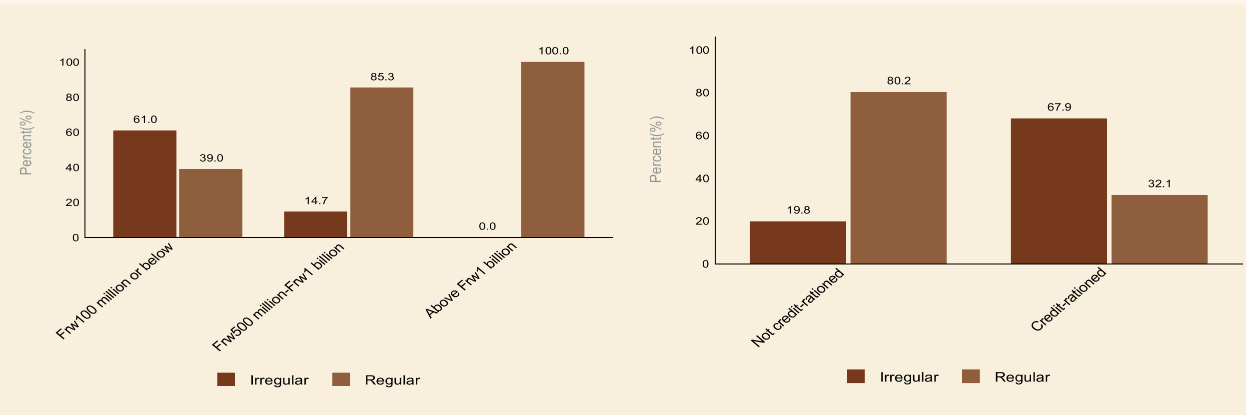
Figure 4: Loan application and rationing among manufacturing firms in Rwanda (2018)



4.1.3. Distribution of credit delinquency by loan amount

Micro-credit accounts were more likely to fall under the NPL category. As reported in Figure 5, loans amounting to less than 100 million francs were 46 percentage points more likely to end up in the NPL category, compared to larger loans. Interestingly, Figure 5 further reports higher scores of NPLs among credit-rationed firms. Loan delinquency among credit-rationed borrowers is found to be 48 percentage points higher, compared to their non-credit-rationed counterparts. This suggests that banks are more likely to ration amount for borrowers with bad credit histories.

Figure 5: Credit repayment performance by loan size and amount rationing by the bank (2018)

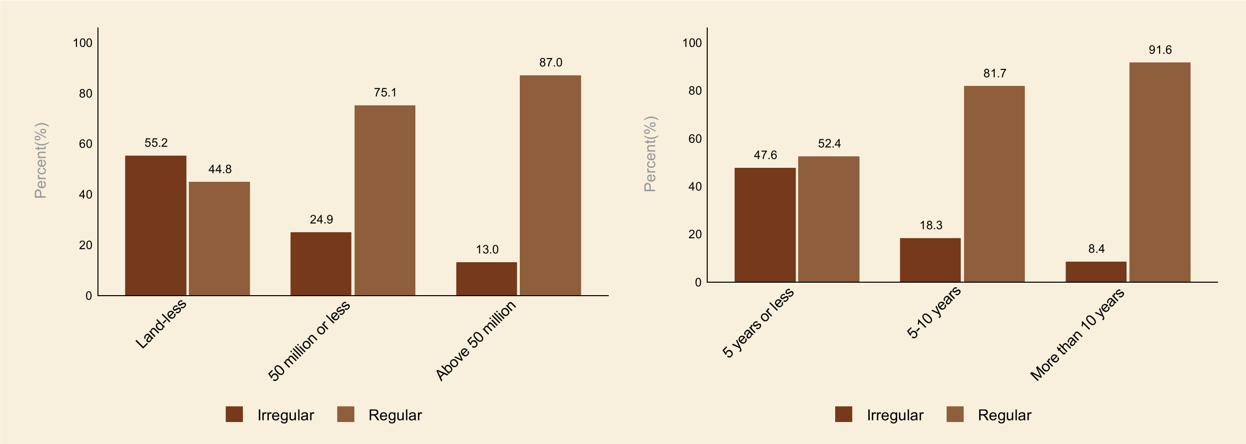


Source: Authors' analysis using survey data

4.1.4. Credit delinquency and collateral value and credit history

As expected, loan delinquency decreases significantly with landholding and years of experience with credit. Incidences of loan delinquency were about 30 percentage points higher among landless firms compared to manufacturers with land properties (see Figure 6). Moreover, the prevalence of NPLs was 29 percentage points higher among manufacturers with less than 5 years of experience with credit. This finding supports credit score as an important indicator based on which lenders can evaluate the creditworthiness of manufacturers.

Figure 6: Loan repayment performance by land value and experience with credit (2018)



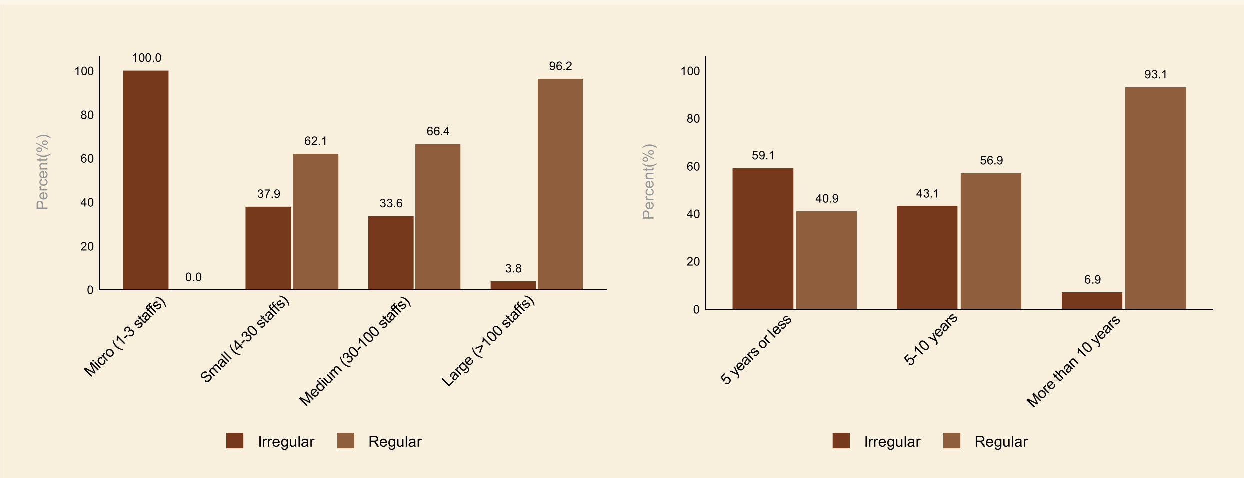
Source: Authors' analysis using survey data

4.1.5. Prevalence of credit delinquency across firm size and age categories

Some firm-specific characteristics turn out to be stark delineators of the propensity to default. Figure 7 reveals that, compared to small and medium enterprises, the incidence of delinquent loan account was 30 percent lower among larger enterprises. Across firm age categories, the prevalence of credit delinquency among ten years and older firms was 36 percentage points lower compared to younger firms. This finding, however, needs to be interpreted with caution, as the size and age effect is not unidirectional.

Bad credit history among loan delinquents could significantly hinder access to external financing, thereby stunting enterprise growth. At the same time, smaller enterprises might lack the necessary managerial capacity to manage external funds, leading to credit delinquency.

Figure 7: Credit delinquency by firm size and age (2018)

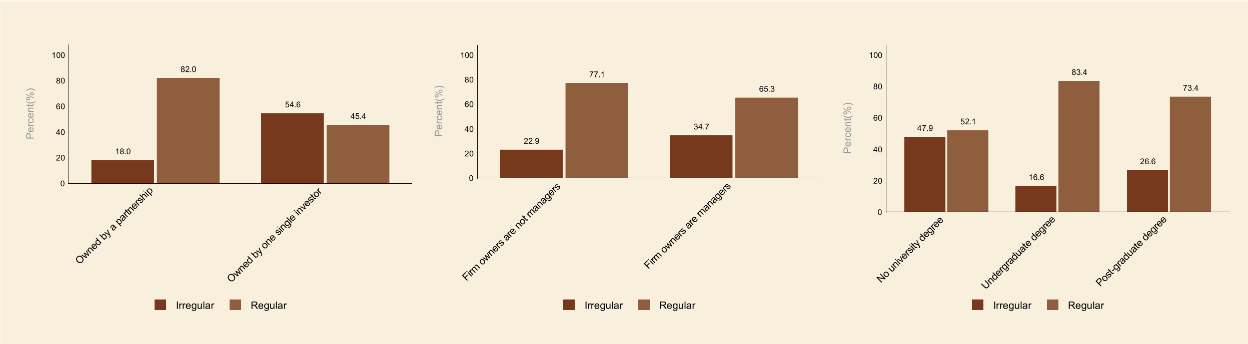


Source: Authors' analysis using data from survey.

4.1.6. Prevalence of loan delinquency across firm ownership and management

Evidence from the survey data shows an unevenly distributed loan delinquency across various aspects of firm management. Notably, Figure 8 shows higher incidences of credit delinquency among firms owned by one single investor. The incidence of delinquent loan account was 36.6 percentage points higher among firms owned by one single entrepreneur.

Figure 8: Loan repayment performance by firm ownership and management (2018)

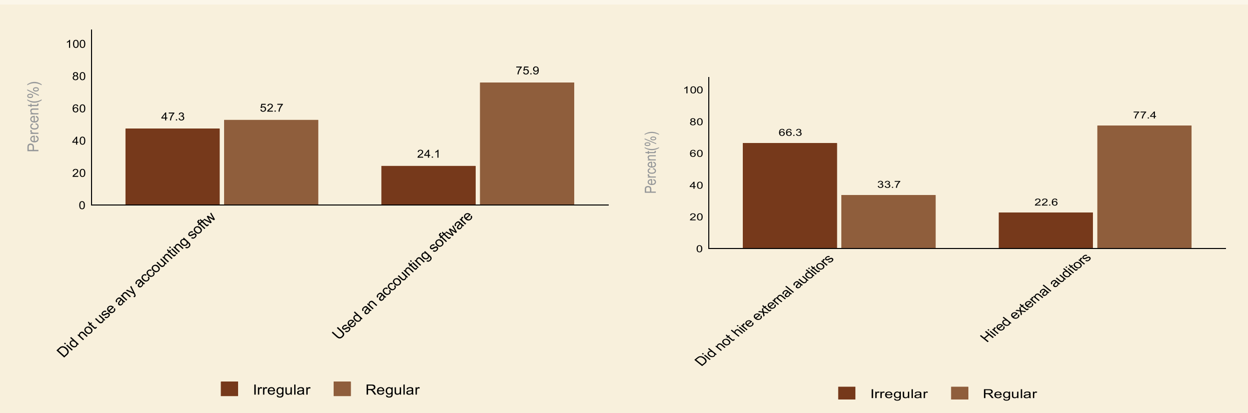


Source: Authors' analysis using survey data

The odds of loan delinquency were higher among firms managed by executives without tertiary education. The prevalence of delinquent loan accounts was 20 percentage points higher among manufacturers without tertiary education compared to their more educated counterparts.

Moreover, analysis of survey data suggests that the likelihood of loan delinquency was higher among financially opaque manufacturers. As plotted in Figure 9, the incidences of delinquent loan account were respectively 23 and 43 percentage points lower among manufacturing firms that used accounting software and those that held externally audited financial reports compared to their financially opaque counterparts. These results point to the importance of information asymmetry for loan performance.

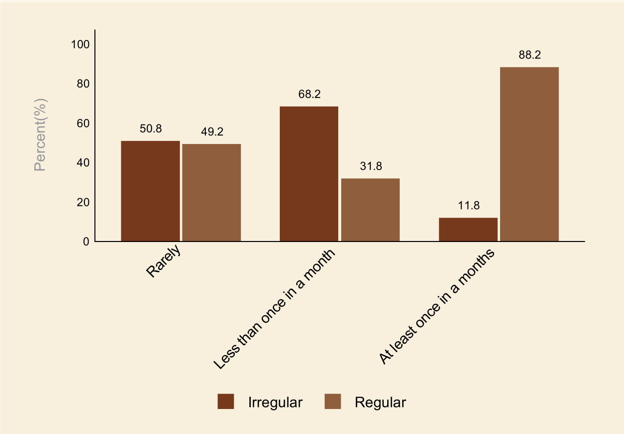
Figure 9: Loan repayment performance by financial transparency (2018)



4.1.7. Loan delinquency and relationship with banks

Bank visits, an indicator of the quality of bank relationship with the borrower, does not significantly alleviate loan delinquency. The results in Figure 10 shows no specific pattern of loan delinquency distribution across categories of bank visits. Nevertheless, credit delinquency score was lower among firms that receive/make at least one bank visit per month.

Figure 10: Credit repayment performance by frequency of bank visit (2018)

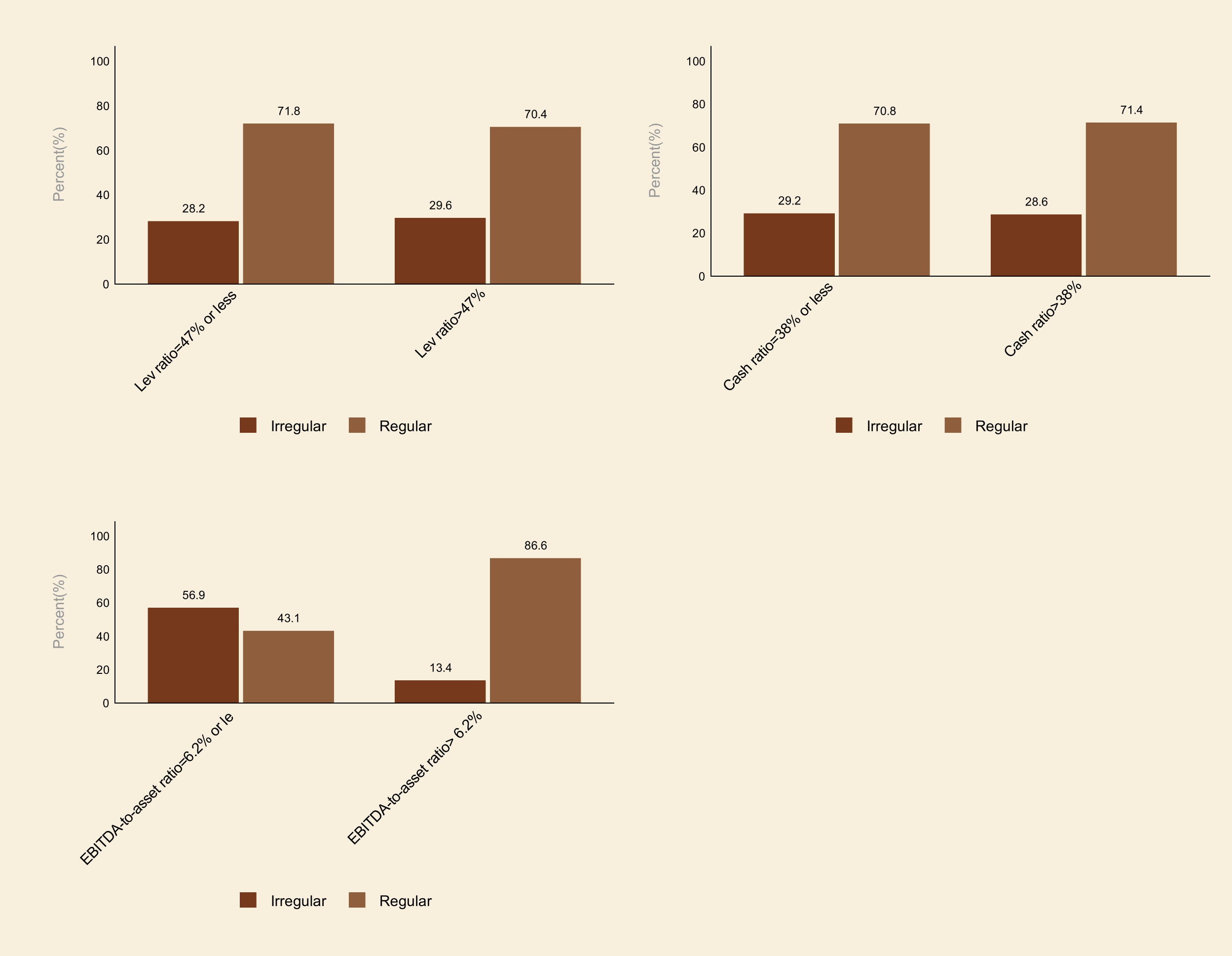


Source: Authors' analysis using survey data

4.1.8. Distribution of loan delinquency across traditional financial ratios

Loan delinquency is evenly distributed across firms in various categories for each traditional financial ratio, except for Profitability (i.e. EBITDA-to-asset ratio). The data plotted in Figure 11 point to relatively higher proportions of delinquent loan accounts among less profitable firms. The incidence of delinquent loan account is 43 percentage points greater among firms in the below-median category compared to their counterparts. This finding suggests that firms that are more efficient in utilizing their assets to generate more profit are also more likely to perform better on their loan accounts.

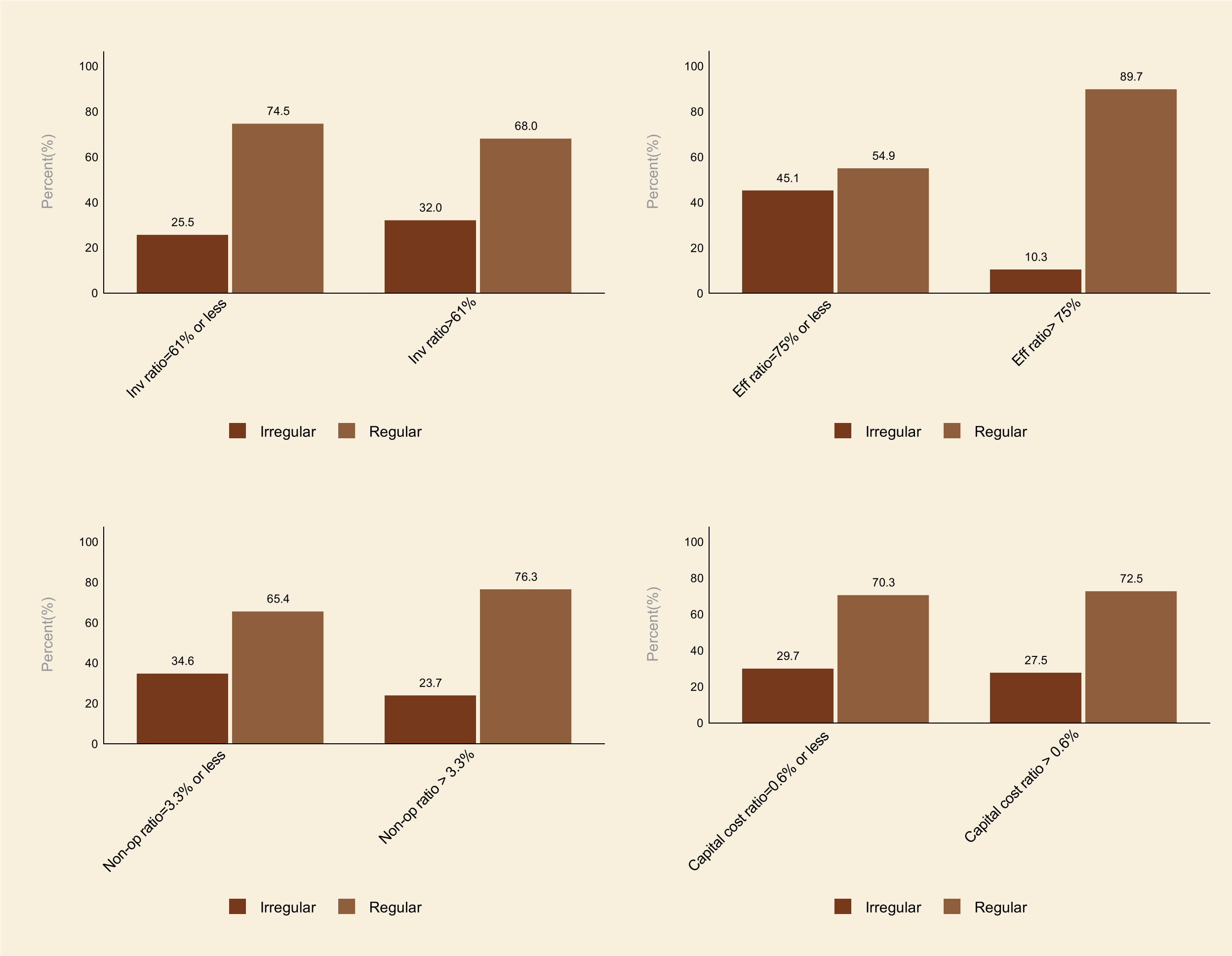
Figure 11: Loan repayment performance by firm liquidity, profitability, leverage and solvency (2018)



4.1.9. Distribution of loan delinquency across financial ratios for moral hazard

Credit delinquency was fairly distributed across various categories of observable indicators used to signal potential moral hazard, except for the financial ratio for effort (i.e. operating costs and expenses over total assets). As shown in Figure 11, the incidences of having a delinquent loan account were disproportionately distributed across firms in the two categories of Effort. In line with the theoretical expectations, the incidence of irregular repayment was nearly 35 percentage points higher among firms in the below-median category, compared to those in the upper efforts category. This result indicates the extent to which the intensity with which managers make efforts to keep the business afloat delineates the odds of running into loan delinquency.

Figure 12: Loan repayment performance by financial ratios for moral hazard (2018)



Source: Authors' analysis using survey data

4.2. Logit estimation results

Table 4 presents the results on two different probit models for estimating the effect of financial distress and moral hazard on the probability of being credit-delinquent. Model 1 was estimated by using traditional financial ratios to predict credit delinquency while controlling for firm-specific characteristics. Model 2 adds to Model 1 the different financial ratios for moral hazard among other independent variables as specified in Section 3. According to various diagnostic tests, Model 2 gives the best fit. This model is the best model according to the -2loglikelihood value and confirmed by Nagelkerke's R-squared. Moreover, the Hosmer–Lemeshow test shows that both models are correctly specified.

The results in Table 4 show that the coefficients of all three traditional financial ratios used in the model have expected signs. In line with the theoretical expectations of the financial distress hypothesis (see Table 1), the coefficients of LIQUIDITY and PROFITABILITY are negative and significant in both models.

The financial ratios for moral hazard also have the expected signs but interestingly, none of them turns out to be significant in either model.

The coefficients of three firm-specific characteristics have the expected signs, expect SINGLEOWNER. In both models, the coefficient of SINGLEOWNER has a positive sign, which is counter-intuitive based on the theoretical expectations of the moral hazard hypothesis. Nonetheless, its coefficient was insignificant in both models. The coefficient of FIRMAGE turns out to be negative and significant, while the coefficient of its quadratic form of firm size (FIRMAGEˆ2) is negative but insignificant in the model. LANDBUILDVAL has a negative bust insignificant coefficient in both models.

4.3. Discussion

The results of all three models show that the traditional financial ratios of liquidity and profitability are the most significant predictors of the probability of credit default. The significantly negative coefficient of LIQUIDITY (i.e. the ratio of cash of total assets) in Models 1&2 suggests that the odds of credit delinquency decrease with the importance of liquidity in a firm's asset. Echoing the firm manager's contention that cash-flow-related issues were among the most important challenges faced when repaying loans (see Figure 4), the finding suggests that the ease with which firm assets can be liquidated can predict the probability of default. The significant effect of liquidity has been reported in other case studies (e.g. Castillo et al., 2017).

The negative and significant effect of PROFITABILITY (i.e. EBITDA-to-total asset ratio) across the two model specifications suggests that the probability of holding a delinquent loan account decreases with business profitability. The revealed negative coefficient is suggestive of the potential financial distress effect. As found by previous empirical studies that explicitly distinguish strategic defaulters from financially distressed defaulters (Asimakopoulos, et al., 2016), defaulters with lower profitability are less likely to be strategic defaulters. The negative effect of PROFITABILITY in this study, therefore, reveals credit delinquency as an outcome of inefficiency in utilizing the firm's resources to generate profits.

The negative and significant coefficient of FIRMAGE suggests that, ceteris paribus, younger firms are more likely to default on their credit accounts than their well-established counterparts. This finding is in line with the previous studies that have unveiled a negative effect of firm experience on the probability of default (Diamond, 1989; Hyytinen & Väänänen , 2006).

Overall, our findings fail to provide clear evidence of moral hazard among loan-delinquent manufacturers in Rwanda. This conclusion echoes the findings of previous studies that have explicitly distinguished moral hazard effects, showing that moral hazard is less prevalent compared to adverse selection (Hyytinen & Väänänen , 2006).

Table 4: Estimation results of logit model of credit delinquency

	Model 1		Model 2	
Independent variables	**Coefficient**	**P>z**	**Coefficient**	**P>z**
Traditional financial ratios				
LIQUIDITY	−87.173	(0.069)	−91.955	(0.060)
PROFITABILITY	−4.204	(0.007)	−3.625	(0.037)
LEVERAGE	0.184	(0.610)	0.199	(0.595)
Financial ratios for moral hazard				
SUBSTITUTION			−681.296	(0.235)
EFFORT			2.913	(0.131)
UNDERINVEST			−1.638	(0.414)
DIVERSION			−2.369	(0.287)
Firm-specific characteristics				
SINGLEOWNER	0.049	(0.197)	0.0010	(0.206)
FIRMAGE	-0.0602	(0.000)	-0.0010	(0.002)
FIRMAGE^2	0.0005	(0.180)	0.0010	(0.178)
LANDBUILDVAL	-0.0028	(0.108)	-0.0011	(0.147)
Goodness of fit				
2loglikelihood	207.878		130.542	
Nagelkerke's R^2	0.119		0.171	
Hosmer and Lemeshow: $\chi 2$	7.224 (0.513)		3.778 (0.877)	

5. SUMMARY, CONCLUSION AND POLICY IMPLICATIONS

Over the last decade, a higher and growing stock of non-performing loans (NPLs) remained among the most challenging aspect of bank asset quality in Rwanda. The manufacturing sector was among the main contributors to the deterioration of bank asset quality. This study investigates the micro-level factors explaining credit default risk among manufacturing firms in Rwanda. The aim is to answer a key policy question: was the surge of NPL from the manufacturing sector in 2018 an outcome of financial distress, or a moral hazard effect (or both)? To this end, the study uses a unique dataset compiled from a cross-sectional survey on access to finance conducted among 122 manufacturing firms in 2018. Based on the previous literature, traditional financial ratios of liquidity, profitability, and leverage were used alongside four financial ratios indicating moral hazard characteristics (asset substitution, low effort, underinvestment, and different use of invested capital) in a logit model.

Overall, the results suggest that the effects of both observable and unobservable indicators of moral hazard characteristics were not significant in the model. They show that the probability of loan delinquency rather decreases with firm profitability and liquidity, which undermines any suspicion of strategic default risk. These effects remain significant even after controlling for firm-specific characteristics.

The findings of this study have several implications in terms of banking regulation and financial stability policy in Rwanda. The regulator (NBR) should review the risk management systems and procedures followed by banks in order to avert future financial instability. In addition to using collateral as an important screening device for customer's credit risk assessment, the findings also point to the value of financial information (such as firm age) and experience for credit risk assessment and management. When assessing creditworthiness, the combined use of financial ratios and non-financial details would lead to a more accurate explanation of current and future loan repayment performance among manufacturing firms.

Financial stability monitoring by NBR should consider developing macro-prudential tools for monitoring and predicting credit default risk in the manufacturing industry and other sectors. The revealed link between credit default risk and financial distress suggests that simple but accurate financial distress prediction models are of critical importance to various stakeholders (including the regulator, banks, and borrowers) as they would provide them with early warnings. BNR should capitalize on its access to a wide range of firm data (i.e. big data) to develop analytics processes and techniques (such as data mining, predictive analytics, and machine learning) for forecasting industry-specific financial performance and link this with the current and future performance of loan portfolios.

REFERENCES

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23(4), pp. 589-609.

Altman, E. I. & Sabato, G., 2007. Modelling credit risk for SMEs: Evidence from the U.S. market. ABACUS A Journal of Accounting, Finance and Business Studies, Volume 43, pp. 332-357.

Asimakopoulos, I., Avramidis, P. K., Malliaropulos, D. & Travlos, N. G., 2016. Moral hazard and strategic default:evidence from Greek corporate loans. Bank of Greece Working Paper.

Bebczuk, R. N. & Bebczuk, R. N., 2003. Asymmetric information in financial markets: introduction and applications. Cambridge University Press.

Castillo, J. A., Mora-Valencia, A. & Perote, J., 2017. Moral hazard and default risk of SMEs with collateralized loans. Finance Research Letters, Volume 26, pp. 95-99.

Ciampi, F., Giannozzi, A., Marzi, G. & Altman , E. I., 2021. Rethinking SME default prediction: a systematic literature review and future perspectives. Scientometrics, Volume In press.

Diamond, D. W., 1989. Reputation acquisition in debt market. Journal of Political Economy, Volume 97, pp. 828-862.

Fidrmuc, J. & Hainz, C., 2010. Default Rates in the Loan Market for SMEs: Evidence from Slovakia. Economic Systems, Volume 34, pp. 133-147.

Giroud, X., Mueller, H. M., Stomper, A. & Westerkamp, A., 2012. Snow and leverage. The Review of Financial Studies, Volume 25, p. 680–710.

Guiso, L., Sapienza, P. & Zingales, L., 2013. The determinants of attitudes toward strategic default on mortgages. The Journal of Finance, Volume 68, pp. 1473-1515.

Hitayezu, P., Kamanzi, C., Manishimwe, C. & Kamugisha, C., 2019. Access to credit among manufacturing firms in Rwanda: An assessment of supply-side challenges. BNR Economic Review, Volume 15, pp. 57-92.

Hyytinen, A. & Väänänen , L., 2006. Where do financial constraints originate from? An empirical analysis of adverse selection and moral hazard in capital markets. Small Business Economics, Volume 27, p. 323–348.

Karuhanga, W. K., Maniraguha, F. & Hakuzwimana, A., 2018. Determinants of non-performing loans in Rwandan banking sector.. BNR Economic Review, Volume 12, pp. 33-59.

Le, C. H. A. & Nguyen, H. L., 2019. Collateral quality and loan default risk: The case of Vietnam. Comparative Economic Studies, Volume 61, p. 103–118.

NBR, 2017. Annual Financial Stability Report 2016-2017, Kigali, Rwanda: National Bank of Rwanda.

NBR, 2020. Annual Financial Stability Report: July 2019 - June 2020, Kigali, Rwanda: National Bank of Rwanda.

Stiglitz, J. & Weiss, A., 1981. Credit rationing in markets with imperfect information.
The American Economic Review, 71(3), pp. 393-410.

World Bank, 2020. Future Drivers of Growth in Rwanda : Innovation, Integration, Agglomeration, and Competition. Washington, DC: World Bank.

MEASURING THE WELFARE COST OF INFLATION IN RWANDA

Josine Uwilingiye[7]
Patrick Hitayezu[8]
Faustin Maniraguha[9]

[7] Senior Lecturer, School of Economics, University of Johannesburg
[8] Director, Research Department, National Bank of Rwanda
[9] Senior Economist, Research Department, National Bank of Rwanda

ABSTRACT

In 2019, the National Bank of Rwanda adopted a price-based monetary policy framework. With price stability as the main objective of monetary policy, the central bank committed to maintaining a medium-term inflation benchmark of 5±3 percent using discretionary changes in its Central Bank Rate (CBR). This paper sets out to investigate the magnitude of the economic cost associated with inflation in order to provide a basis for judging the desirability of the inflation benchmark. Using quarterly observations from 2006:3 to 2019:4, two alternative money-demand function specifications are estimated, namely the double log and the semi-log models. The Dynamic Ordinary Least Squares (DOLS) and autoregressive distributed lad (ARDL) procedures are used to obtain cointegrating regressions for both specifications. Our diagnostic tests suggest that, compared to the semi-log specification, the double log demand function performs better on the Rwanda data. Estimates of the double-log specification from the ARDL procedure turn out to have more predictive power and suggest a welfare cost ranging from 0.5077 to 0.8537 percent of annual real GDP over the medium-term inflation benchmark band. Compared to other economies in the region (such as Kenya), these estimates are relatively higher, which supports the case for continued pursuance of price stability as the primary goal for monetary policy in Rwanda.

Keywords: inflation, money demand, welfare, Rwanda
JEL classification: E31, E41

1. INTRODUCTION

Over the last decade, the transition to inflation targeting among central banks has become a global phenomenon. Worldwide, most central banks have adopted full-fledged inflation targeting regimes or transitional arrangements, and in line with the prescriptions of new Keynesian economics, price stability has increasingly become the primary objective of monetary policy (Laurens, et al., 2015). Inspirited by this global consensus as well as other regional aspirations such as the East African Monetary Union (EAMU), National Bank of Rwanda (NBR) has embarked on a journey to modernize monetary policy. In 2019, NBR adopted a price-based framework and committed to maintaining a medium-term inflation benchmark of 5± 3 percent using discretionary changes in the central bank rate (CBR).

The medium-term inflation benchmark of NBR was set in a bid to harmonize with key macroeconomic convergence criteria underlying the establishment of the East African Monetary Union (EAMU), which includes inter alia a headline inflation ceiling of 8 percent. This "common" inflation criterion for the EAMU was set based on various determinants of inflation target, including past inflation, economic growth, business cycles fluctuations, wage and price rigidities, measurement error, monetary policy flexibility, zero interest rate bound, and welfare cost of inflation (Horvath & Mateju, 2011). Nevertheless, case studies investigating the aforementioned determinants of the medium-term inflation objective to inform the inflation criterion for individual countries and the East African Community (EAC) as a whole remain scanty.

The scarcity of this strand of empirical research is even more pronounced in Rwanda. To date, only the economic growth criterion of the inflation benchmark has received attention among economists in Rwanda. For example, a recent study by Gichondo et al. (2018) showed an inverted U-shaped relationship between GDP growth and inflation with an optimal point at around 5.9% inflation rate, which empirically underpinned the benchmark band. To the best of the authors' knowledge, no study has investigated the magnitude of the cost of inflation to judge the desirability of the 5±3 percent headline inflation benchmark for monetary policy in Rwanda.

In other countries of the East African Community (EAC) region, there have been a few attempts to investigate the welfare cost of inflation. In Kenya, for example, a study conducted by Ikikii (2017) reported a trivial welfare cost of inflation equivalent to 0.041 and 0.103 percent of annual real GDP for an inflation band of 5%± 2.5%. Although Kenya and Rwanda share some common economic features, the welfare cost of inflation could be heterogeneous across these economies, given potential differences in interest elasticity of money demand. A recent empirical study by Dunne and Kasekende (2018) revealed that Kenya and Tanzania were among Sub-Saharan countries that had recorded higher growth rates of financial innovation and significantly lower demand for cash, which weakened the relationship between inflation and income. This assertion is in line with recent evidence suggesting that inflation imposes a disproportionately higher welfare cost to countries at earlier stages of financial innovation (Mogliani & Urga, 2018;

Berentsen, et al., 2015; Cao, et al., 2020). It is, therefore, necessary to investigate the orders of magnitude of the economic losses associated with the negative consequences of the cost imposed by inflation on a case-by-case (and where necessary period-by-period) basis to inform individual central banks in the region on the desirable levels of inflation benchmark.

Rwanda presents an appealing case for investigating the magnitude of the cost of inflation. Despite recent developments in financial innovations, various reports by NBR show that money holding by households and firms remains a persistent aspect of the financial economy. Available data on money supply suggests that M1, an indicator of the country's money supply used as a medium of exchange, has been growing rapidly. Recent estimates suggest that the ratio of M1 to GDP has been growing on an average of 1.1 percent per year. For the sample period of 2006 to 2019, a 1 percent increase in nominal GDP led to around 1.23 percent increase in M1, ceteris paribus.

Against this backdrop, this study attempts to empirically estimate the welfare cost of inflation to shed light on the desirability of 5±3 percent as the medium-term inflation benchmark for Rwanda. Although empirical literature has vindicated the superiority of general equilibrium frameworks in providing a more accurate account of the welfare cost of inflation (Dotsey & Ireland, 1996; Shah, et al., 2019), this study applies the approaches of the traditional partial equilibrium technique of consumer surplus Bailey (1956) for simplicity purposes. We estimate two alternative money-demand function specifications à la Lucas (2000)[10] , namely the double log model of demand for money of Meltzer (1963) and the semi-log model of Cagan (1956) using quarterly observations from 2006:3 to 2019:4. These two functional forms with different curvatures at lower nominal interest rates Ireland (2009) have been widely tested and proven to fit differently money demand functions in several developing economies[11].

Inspired by recent studies such as Ikikii (2017) and Shah et al. (2019), our contribution focuses on the selection of the best model for Rwanda. Based on both Dynamic Ordinary Least Squares (DOLS) and Autoregressive Distributed Lag (ARDL) estimates, we find that estimates of welfare loss greatly depend on the functional specification of the money-demand curve. We show that the welfare cost of inflation under a double log model of demand for money generates not only more reliable estimates but also yields a substantially higher welfare loss of GDP over the benchmark inflation band of 5±3 percent, compared to the semi-log model. These findings are supported by both the DOLS and the ARDL procedures, although ARDL estimates (ranging from 0.5077 to 0.8537 percent of annual real GDP over the medium-term benchmark band) turn out to have more predictive power.

The remainder of this paper is organized into five sections. Section 2 presents some key stylized facts about inflation in Rwanda. Section 3 reviews the theoretical and empirical literature on the welfare cost of inflation. Section 4 outlines the methodology, while Section 5 reports and discusses the findings. The last section summarizes the key findings and outlines key policy implications.

[10] Based on a simple framework of perfectly anticipated inflation with no other distortions.
[11] See (Gupta & Uwilingiye, 2009) (Mushtaq, et al., 2012) (Ikikii, 2017) (Shah, et al., 2019)

2. OVERVIEW OF RECENT INFLATION DYNAMICS IN RWANDA

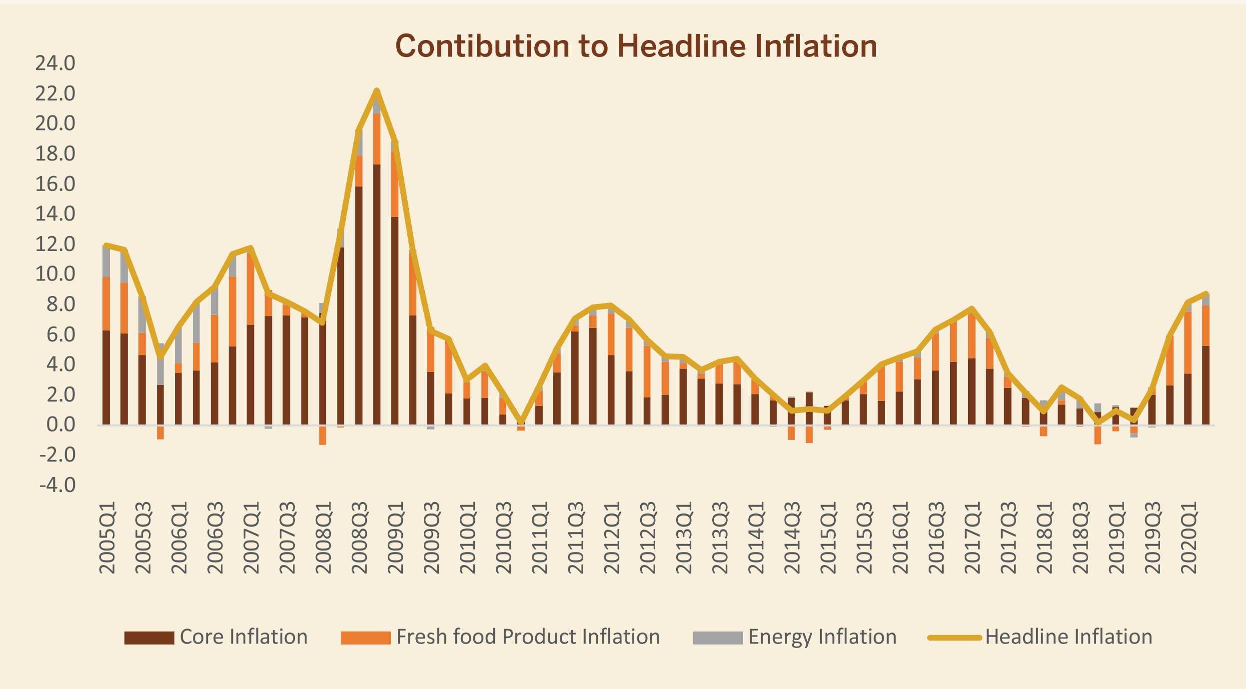
Under its mandate of monetary policy[12] , NBR seeks to promote price stability in Rwanda. The central bank uses data on inflation developments based on a consumer price index (CPI) that captures the weighted average of prices of a basket of consumer goods and services. The data are released on a monthly basis by the National Institute of Statistics for Rwanda (NISR).

As Figure 1 shows, the inflation rate fluctuated substantially in the pre-global financial crisis (GFC) era. This era was often marked with exogenous supply shocks related to oil price shocks. These shocks were transmitted to consumer prices via direct channels (i.e. energy inflation) or indirect processed (cost-push inflation).

After the GFC, inflation generally stabilized, as quarterly rates remained confined within the lower and upper band of the benchmark. During this period, episodes of higher inflations have been associated with developments in agriculture and episodes of galloping exchange rate depreciation (e.g. 2016-2017). Experiences with supply shocks from the agricultural sector, mainly in the form of droughts, floods, and landslides, have often underlain persistent upticks in inflation. The prolonged droughts of 2015-2016, as an example, led to an uptick of inflation from just under 2.0 percent in 2015Q2 to 3.5 percent in 2017Q3, peaking at 7.7 percent in 2017Q1.

Overall, headline CPI inflation has averaged 3.9 percent since 2010, which is below the benchmark of 5.0 percent. Where inflation rate has crossed outside the benchmark band, this has often been a result of a base effect of inflation in the corresponding period of the previous year.

Figure 1: Contributors to headline inflation in Rwanda.



Source: National Institute of Statistics for Rwanda (NISR)

[12] Accorded by Law N°48/2017 of 23/09/2017 governing the NBR (often referred to as Central Bank Law),

3. LITERATURE REVIEW

3.1. Costs of inflation

One of the longstanding questions in contemporary economic literature has been the cost of monetary policies of positive (but low) inflation, compared to the disinflation implied in Milton Friedman's optimal monetary policy of zero nominal interest rate (Ireland, 2009). It is the understanding of such costs that justifies the desirability of price stability as the primary goal of monetary policy (Dotsey & Ireland, 1996; Driffill, et al., 1990). Indeed, Feldstein (1979) argues that since a permanent increase in inflation buys only a temporary reduction in unemployment, expansionary fiscal policy decision-making requires a comparison between the present value of perpetual welfare cost of inflation with welfare gains of reduction in unemployment.

Therefore, Macroeconomic models should tell which optimal levels of monetary and fiscal policy are needed to achieve the preferred combination of price stability and unemployment objectives (Leigh-Pemberton, 1992).

Economists have therefore devoted efforts to identify and discuss the costs associated with positive inflation rates. The economic literature divides the costs into two categories, namely costs due to fully anticipated inflation and the costs due to unanticipated inflation. The costs of anticipated inflation, the focus of this study, are clearly articulated in (Balls & O'Donnell, 2002). These include:

1) The costs of economizing on real money balances, also known as shoe-leather cost Pakko (1998): capture the opportunity cost of time and effort that people expend by holding less cash in order to reduce the inflation tax that they pay on cash holdings when there is high inflation;

2) Inflation acting as a tax in a less-than perfectly indexed tax system: holding headline tax constant, effective tax rates increase with inflation in a tax system that operates in nominal terms (Bakhshi, et al., 1999).

3) The costs of constantly revising price lists (also known as menu costs): these are costs that arise out of the need to quote new prices, as sellers change menus, reprogram their vending machines, re-advertise prices, redraft contract terms, etc.

4) The problems of 'front-end loading' of the real burden of servicing nominal debt contracts: the real value of fixed nominal debt is higher early in the contract but declines with inflation.

Leigh-Pemberton (1992) argues that the costs outlined above are usually well known and predictable, and societies have made efforts to minimize them (e.g. information technology to reduce the shoe-leather and menu costs). Based on Fisher and Keynesian views, the author argues that most costs of inflation arise from its uncertainty. Though the latter category is beyond the scope of the present study, it is worth mentioning that the costs of unanticipated inflation include:

1) The welfare loss associated with distributional effects of unanticipated changes in real interest rate: by reducing the real burden of debt and damaging creditors, an unanticipated decrease in real interest rates alters the distribution of income and wealth.

2) Distortionary effects related to inefficient allocation of resources: with the typically higher and often skewed price variability observed at high levels of inflation, relative price signals can be confused by aggregate inflation, making it hard for investors to optimally allocate their resources to different productive activities. Also, savers may be discouraged by the uncertainty in real interest rate.

3) The costs of efforts directed towards anticipating inflation and offsetting its unwelcome effects: when inflation cannot be anticipated with a fair degree of certainty, firms are more willing to engage in production activities with short-term payoffs rather than long-term contracts, and money savers and lenders demand a risk premium, which increases the real cost of funds for borrowers.

3.2. Theory and measurement

The roots of the cost of anticipated inflation can be traced back in the basic feature of the shopping time model, i.e. carrying money reduces the time required for conducting transactions, which is, in essence, the convenience or benefit of holding cash. When facing inflation, individuals and firms would have to manage the trade-off between the convenience of using money to conduct transactions and the opportunity cost of holding non-interest-bearing money – i.e. nominal interest rate – which is the basic feature of the Baumol-Tobin's inventory approach to money demand. Therefore, the welfare cost of inflation can be viewed as a general equilibrium version of the Baumol-Tobin's inventory approach to money demand (Pakko, 1998).

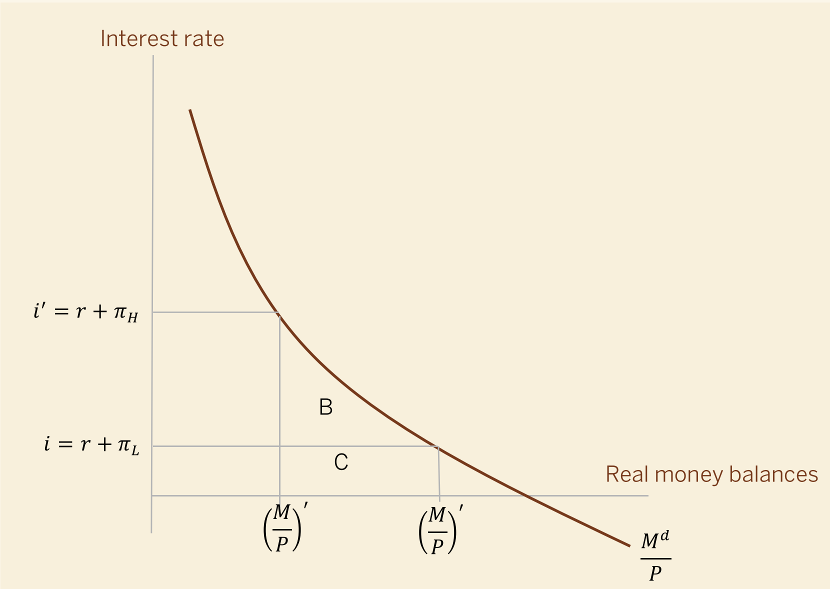
Two approaches to measuring the welfare cost of anticipated inflation can be found in the literature, namely the consumer surplus approach by Bailey (1956) and the compensating variation approach (Lucas, 2000). Bailey (1956)'s original idea is based on the view that the welfare cost of anticipated inflation operates as a tax. As elaborated in Section 3.1, there are two specific channels through which inflation operates as a tax (Bakhshi, et al., 1999): (i) through the effect of inflation of 'effective tax rate' under a less than fully indexed tax system; and (ii) as a direct tax on holdings of money balances (as inflation erodes the real value of money balances, inducing money holders to deposit their balances with banks, a costly choice in a welfare sense) (Bakhshi, et al., 1999).

With such a consideration, Bailey (1956) argues that the evaluation of cost of anticipated inflation is reducibly a standard application of welfare effect of any distortion.

Based on a standard Fisher equation which presents real interest rate (r_t) and inflation expectation (π_e^t)as two components of nominal interest rate (i_t), (i.e. $i_t=r_t+\pi_e^t$), Bailey (1956) draws a liquidity preference as a function of nominal interest rate and shows that welfare losses of increased inflation rate would simply capture the area of unsatisfied demand beneath the demand curve for real money balances. The area marked as BC (trapezium) in Figure 2 measures the amount of consumer surplus forgone as a result of inflation, which corresponds to the deadweight loss of welfare, i.e. loss in consumer surplus minus seignorage revenue.

The compensating variation approach to measuring the welfare cost of inflation proposed by Lucas (2000) built on the Money-In-the-Utility (MIU) model to define the welfare cost of a nominal interest rate r, W(r) as the income compensation needed to leave the household indifferent between living in a steady state with an interest rate constant at r and an otherwise identical steady-state with the interest rate of zero[13].

Figure 2: Money demand curve and welfare cost of inflation



Source: Adapted from Bailey (1956) and Pakko (1998)

13 See illustration in (Gupta & Uwilingiye, 2009)

3.3. Brief overview of the empirical literature

Quantifying the welfare cost of anticipated inflation has been the subject of various empirical studies. The starting point of the discourse on the significance of welfare costs of inflation was the growing evidence of a weak relationship between inflation and economic growth. Cross-country evidence reported by Barro (2013), as an example, suggested a very little relationship between inflation and annual GDP growth rates for inflation rates below 10 percent, despite stark negative relationships above the 10 percent inflation mark.

Early case studies that have primarily investigated the welfare cost of inflation found surprisingly lower costs of inflation, which somehow supports Barro (2013)'s findings. In the US, the focus of the majority of studies in this strand, Lucas (2000) provides a general equilibrium justification for Bailey's consumer surplus approach by proposing a compensating variation approach, and based on a double log money demand function (i.e. constant elasticity of -0.5 percent) with money defined as M1, he reports that reducing the interest rate from 3% to 0% yields a benefit equivalent to an increase in real output of about 0.009 (or 0.9%).

Lucas's (2000) approach has been widely adopted in various case studies around the world. Serletis and Yavari (2004), using data from post-world war II US and Canada in a double-log money demand function, reported much lower welfare cost estimates, showing that reducing the interest rate from 3% to 0% would yield a benefit equivalent to 0.0018 (less than 0.1 percent) of real income. In other parts of the region, Chadha et al. (1998) used data from the UK economy in a semi-log demand function and estimated that 0.22% of GNP is gained by moving from 6% to 2% nominal interest rates. In Africa, exemplary empirical studies are reported in Table 1 below.

Figure 2: Money demand curve and welfare cost of inflation

Study	Country	Approach	Inflation comparison	Welfare cost (percent of GDP)
Gupta & Uwilingiye (2009)	South Africa	Consumer surplus and Compensanting variation	3% to 6%	0.16 to 0.36
Ikikii (2017)	Kenya	Consumer surplus	2.5% to 7.5%	0.041 and 0.103
Makochekanwa (2008)	Zimbabwe	Consumer surplus	10% to 300%	0.9 and 23.4

Source: Authors.

Concerned by the trivialization of the welfare cost of inflation in the empirical literature based on the consumer surplus approach, economists have devoted extra efforts to the empirical measurements. In an attempt to show how the partial equilibrium approach can seriously underestimate the true cost of inflation, Dotsey & Ireland (1996) proposed a general equilibrium model that would capture the cost related to the inefficient substitution of out-of-market activity and into leisure, as well as the devotion of productive time to activities that enable to economize on cash balances. The authors revealed that a sustained 4 percent inflation in the US costs the economy the equivalent of over 1 percent of output per year, with money defined as M1. Various empirical studies, including Gupta & Uwilingiye (2010) and Shah et al. (2019), have adopted this approach. Gupta & Uwilingiye (2010) for example, used Dotsey & Ireland, (1996)'s approach to estimate a welfare cost ranging from 0.70 to 1.33 for an inflation target of 3% to 6% in South Africa.

Other studies have focuses on the role of financial innovation that generally reduces the reliance on cash money to make transactions. Indeed, the findings of early studies from developing countries from various developing countries indicate that the costs of failing to innovate in the financial sector increases as the inflation rate rises (Arrau, et al., 1995). Evidence from 35 sub-Saharan African countries provided by Dunne and Kasekende (2018) unveiled a negative relationship between financial innovation and money demand both in the short-run and long-run, showing that countries that had recorded higher growth rates of financial innovation (as proxied by the ratio of M2/M1) such as Kenya and Tanzania had significantly lower demand for cash. The study showed that when financial innovation is included in the model, the coefficient of inflation appears to be slightly lower.

Based on such insights, empirical studies applying the aforementioned Baumol-Tobin's cash inventory model to welfare cost estimation showed that increased diffusion and penetration of financial innovation (e.g. bank branches and ATM terminals) is associated with lower interest rate elasticity of money demand and welfare cost of inflation. In Italy, for example, Alvarez and Lippi (2009) estimated the welfare cost of inflation using individual household data and showed a 40-percent smaller welfare loss, a reduction that was partly attributable to advances in the ATM withdraw technology. In Canada, Cao et al. (2020) using household survey data showed that across households, money demand increases with age (3 times higher among households aged between 76 and 85 than among households aged 35 or younger) and decreases with household's consumption level (3 times lower among households in the upper-quintile compared to households in lowest quintile), an effect mediated by differential access and use of financial innovations.

3.4. Theoretical foundation

The estimates of welfare cost of inflation rely mainly on money demand specifications. Several studies have evaluated the welfare cost of inflation based money demand specification,

including those of Lucas (2000), Serletis and Yavari, (2004), Serletis and Yavari (2005), Serletis and Yavari, (2007), Gupta & Uwilingiye, (2008), Ireland (2009) and Mushtaq et al. (2012).
Departing from Keynes's liquidity preference theory, the money demand function can be expressed as:

$$\frac{M_t}{P_t} = L(i_t, y_t) \tag{1}$$

Where M_t is nominal money balances P_t is the price level, y_1 is real income and i , represents nominal interest rate.
Assuming that in the long run, L(i,y) function takes the form of:

$$L(i,y) = \phi(i), y \tag{2}$$

the money demand function can be written as $m=\phi(i),y$ where $m=M_t / P_t$ denotes real money balances. Equivalently, it can be expressed as $m/y=\phi(i)$[14] which gives the demand for real money balances per unit of income as a function of the nominal interest rate. Based on Keynes money demand specifications, the two prominent money demand specifications are the double log specification with infinite elasticity at the zero interest rate Meltzer (1963) and the semi-log model elasticity, Cagan (1956) characterized by a time switch from moderate to relatively high semi elasticity at near-zero rates. Meltzer –type money demand for real balances increases without reaching the satiation point as the interest rate approaches zero, while for the Cagan-type semi-log demand model, the real money balances converge to zero as the interest elasticity increases and reaches the satiation point as interest approaches zero[15]. According to Marty (1999), the double log model performs well in times of moderate inflation while the semi-log model fits the data well in the hyperinflation period or with the rate of interest close to zero . Lucas (2000) was the first to derive welfare cost estimates using two money demand specifications in the United States. The double log model instigated by Meltzer (1963) relates the natural logarithms of m (the ratio of nominal money balance to nominal income) as a function of the natural logarithm of short term nominal interest rate, r. This relation can explicitly be written as:

$$\ln(m) = \ln(A) - \eta \ln(r) \tag{3}$$

Where $A>0$ is an intercept, and the coefficient $\eta>0$ represents the absolute value of interest elasticity of money demand. Alternatively, Cagan (1956)'s semi-log model expresses m (the ratio of nominal money balance to income) as a function of short-term interest rate, r which can be written as:

$$\ln(m) = \ln(B) - \xi r \tag{4}$$

[14] Assuming that the real money demand is proportional to income, the unitary income elasticity restriction is imposed and real money balances is specified as function of nominal interest rate.

[15] Bailey (1956) and Friedman (1969) studies use a semi log model on the grounds that semi log model perform better during the period of higher inflation.

Where $B>0$ is an intercept and $\xi>0$ measures the absolute value of the semi-elasticity of money demand with respect to the interest rate. Lucas (2000) shows that at a low interest rate, the two specifications of money demand have very different implications for the welfare cost of inflation when the central bank moves from Friedman (1969)'s rule of zero to a positive short-term interest rate[16] . Therefore, identifying a more precise functional form of money demand is of the utmost importance in evaluating the monetary policy.

Using the baseline welfare cost estimates outlined in Bailey (1956), Lucas (2000) transforms the parameters in (3) and (4) into welfare cost estimates. According to Bailey (1956), the welfare cost of inflation can be measured as the area under the inverse money-demand curve (the "consumer surplus") gained by reducing the interest rate to zero from its existing value $r>0$. Thus, for an estimated money demand function given by $m(r)$ and with $\psi(r)$ inverse demand function, the welfare cost of inflation is calculated as follows:

$$w(r)=\int_{m(r)}^{m(0)}\psi\left(x\right)dx=\int_{0}^{r}m\left(x\right)dx-rm\left(r\right) \tag{5}$$

From the above equation, the welfare cost of inflation is obtained by subtracting the seignorage revenue from the consumer surplus. Building on Bailey (1956)'s consumer surplus approach, Lucas (2000) shows that the double log specification of the money-demand function in Eq.3 corresponds to the level $m(r)=A.r^{-\eta}$. From the level formula, it is evident that as short term interest rates approach zero, the money-income ratio becomes very large, thus, the welfare cost of inflation as a fraction of GDP can be obtained using the following equation:

$$w(r)=A\left(\frac{\eta}{1-\eta}\right)r^{1-\eta} \tag{6}$$

For semi-log specification in Eq. 4, the level of has a finite satiation at point B (the intercept). Hence, the welfare cost of inflation as a fraction of GDP can be calculated using the following equation:

$$w(r)=\frac{B}{\xi}\left[1-\left(1+\xi r\right)e^{-\xi r}\right] \tag{7}$$

Note that Bailey (1956) and Lucas (2000) welfare cost estimate is based on the one-dimensional formula which does not into consideration possible trade-off between less liquid interest-bearing money and more-liquid noninterest-bearing. Cysne and Turchick (2010) indicate that such exclusion may result in overestimation or underestimation of the true welfare cost. However, since the Rwandan capital market is in its infant stage of development, we expect to obtain reliable welfare cost estimates based on the one-dimensional welfare cost estimates arising from non-interest-bearing money which is more liquid.

[16] Lucas (2000)

4. DATA AND METHODOLOGY

4.1. Data

To obtain the welfare cost of inflation estimates using both money demand specification in Eq. 3 and Eq. 4, monetary aggregate M1[17] is used as a measure of nominal money balances and the 91-day Treasury bill rate (tbill), a short term interest rate, represents the opportunity cost of holding money. Both series are monthly time series data spanning June 2006 to December 2019, and they are compiled by the National Bank of Rwanda. Moreover, for the same period, quarterly time series data on nominal GDP as a proxy of nominal income, and GDP deflator as a measure of nominal prices are extracted from the National Institute of Statistics for Rwanda (NISR).

However, since the data on nominal GDP and GDP deflator are only available on quarterly basis, we aggregated nominal money balances, M1, and the 91 Treasury bill rates from monthly to quarterly frequencies to have comparable series. Note that the aggregation of the time series data depends on whether the variable is flow or stock (Marcellino, 1999). For monetary aggregate M1, which is a stock variable, the aggregation is done using a systematic sampling aggregation method, while temporal aggregation is used for 91 Treasury bill rate as it is a flow variable. The money income ratios (nm1r_sa) in Eq.3 and Eq.4 are obtained by dividing the monetary aggregate M1 to nominal GDP.

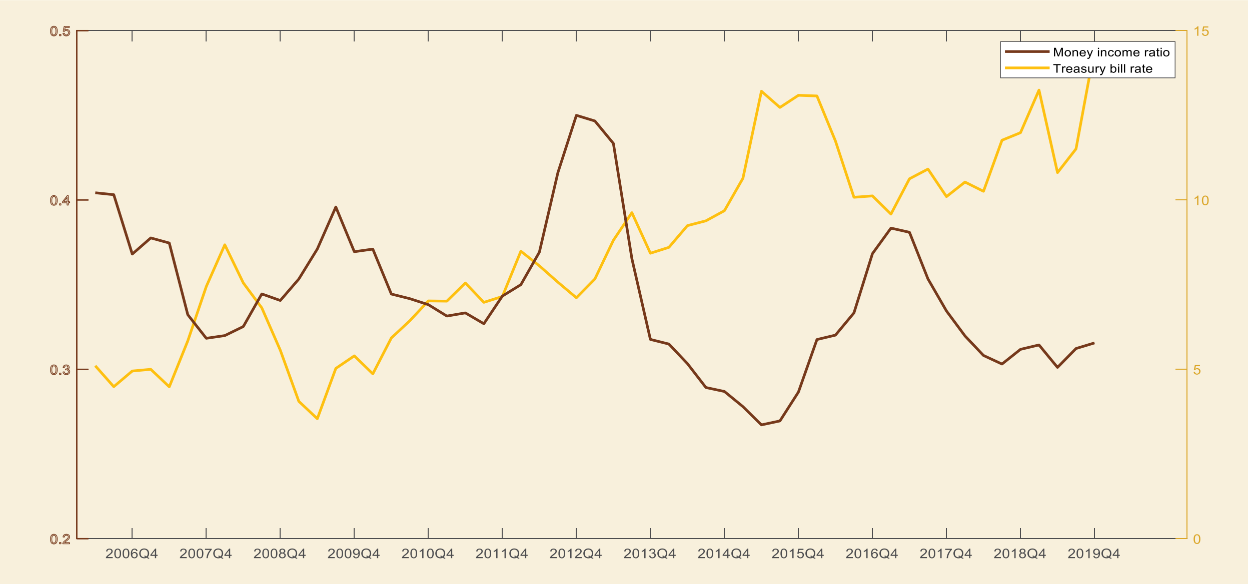
Moreover, both nm1r_sa and tbill are transformed to their natural logarithm lnm1r_sa and lntbill, respectively to estimate the double-log and semi-log money demand specifications. Except for the 91 treasury bill rate series, we employ the TRAMO/SEATS, ARIMA model-based seasonal adjustment method developed by Gómez and Maravall (1996) to deseasonalize the series used in this study.

The visual inspection of the plots of both money income ratio and short-term nominal interest rate is depicted in Figure 3 below.

Looking at the plots, there is a clear indication that the two series move in a different direction over the sample period, 2006q2-2019q4. The nominal interest rate and money income ratio are represented by the red and blue lines, respectively. The nominal interest rate has been varying between 5% to 10%, except for the period around 2012q1 to 2013q1 and 2014q1 to 2015q4.

[17] M1, currency in circulation plus deposit with commercial banks is used as it encompasses to a large extent the transaction demand for money.

Figure 3: Nominal interest rate and money income ratio



Source: Authors

Moreover, the visual inspection shows that that the two series tend to winder away, which indicated that the two series are non-stationary. One of the important aspects when dealing with time series data is to analyze univariate characteristics of each series to avoid spurious regression (Granger & Newbold, 1974). Hereof, Elliott et al. (1996), Dickey-Fuller (DF)-GLS, and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) unit roots tests are used to test the univariate characteristic of lnm1r_sa, lntbill and tbill. As it can be seen from Table A1 (Appendix) all the Lntbr and Tbr variables were found to be stationary while Lnrm1-sa is integrated of order one according to DF-GLS and KPPS. Moreover, as we suspected the presence of structural breaks in the series, for robustness, we also performed unit-roots for structural breaks. We find all the variables to be integrated of order one. The results are reported in Table A2 (Appendix).

4.2. Methodology

The main objective of the study is to compute the welfare cost of inflation in Rwanda using the estimated coefficients emanating from the two money demand specifications Eq.3 and Eq.4 as proposed by Lucas (2000). To obtain the cointegrating vectors we apply single equation cointegration estimators derived from autoregressive distributed lag (ADL) models.

Particularly, we employ Dynamic OLS (DOLS) by Stock and Watson (1993), Saikkonen (1991), and Pesaran et al. (2001) autoregressive distributed lag (ARDL) bound test for cointegration. DOLS methodology approach has been used in a number of studies to estimate money demand model with a single equation cointegration (See, Stock & Watson,1993; Ireland, 2009; Calza & Zaghini, 2011; and Watanabe & Yabu, 2018 among others).

According to Stock & Watson (1993) and Hamilton (1994), in the presence of I(1) variable with single cointegration, the dynamic OLS estimates are asymptotically efficient and asymptotically equivalent to maximum likelihood estimates obtained using the Johansen methodology approach. In this study, since we only have two variables in each money demand specification model, we expect to have at most one cointegrating relation. Hence, DOLS approach is suitable to estimate money demand function.
The Dynamic OLS can specified as follows:

$$Y_t = \alpha + \beta X_t + \sum_{i=-k}^{i=k} \varphi_i \Delta X_{t+i} + u_t \tag{8}$$

Where Y_t represent the dependent variable *lnrm1* and X_t is explanatory variable (either *lntbr* or *tbr*) , and β is the long-run elasticity η or semi elasticity ξ . φ'are coefficient of leads and lags differences of the I(1) tbr and lntbr regressors. The inclusion of lead and lag in Dynamic OLS specification produce an asymptotically efficient estimator that eliminates the reverse causality in the cointegrating system and it may also resolve the problem of autocorrelation. However, if autocorrelation persists, the regression coefficient will remain unbiased but the variance of the regression coefficient will be biased.

Alternatively, to test the robustness of the results obtained using DOLS, we use autoregressive distributed lag (ARDL) bound test for cointegration (Pesaran, et al., 2001). The advantage of ARDL is that its model specification allows lagged dependent variables, current and lagged regressors to be included in the model as regressors. Unlike the residuals obtained from DOLS which tend to be serially correlated Panopoulou and Pittis (2004), the inclusion of lagged dependent variable will tend to produce the uncorrelated residuals (Keele & Kelly, 2006).

Furthermore, for the small sample datasets, the ARDL results are more robust compared to Johansen (1991) methodology approach, Halicioglu (2007) and the ARDL model generates unbiased covariates of the long-run equation (Harris & Sollis, 2003).
The ARDL (p,q) model is specified as follow:

$$Y_t = c_0 + c_1 T + \sum_{i=1}^{p} \phi_1 Y_{t-i} + \sum_{i=0}^{q} \beta_i X_{t-i} + \delta W_t + u_t \tag{9}$$

Where $p \geq 1$ is the number of y_t and $q \geq 0$ is the number of lags for X_t

Following Pesaran et al. (2001), the error correction representation of the ARDL is expressed as follows:

$$\Delta(Y_t) = c_0 + c_1 T + \alpha(Y_{t-1} - \theta X_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta Y_{t-i} + \sum_{i=0}^{q-1} \psi'_{xi} \Delta Y_{t-i} + \delta w_t + u_t \tag{10}$$

Where $\theta = \frac{\sum_{j=0}^{q} \beta_j}{\alpha}$, is the dependent variable($lnrm1\_sa$) , is explanatory variable (either $lntbr$ or tbr) , and is the long-run elasticity, c_0 represents the intercept, T is trend, and W_t are fixed exogenous variables without lags.

5. EMPIRICAL FINDINGS AND DISCUSSION

5.1. Elasticities estimates

The estimated DOLS coefficients are illustrated in Table 2. For the dynamic specification of DOLS, the optimal number lead and lag of change in the interest rate for both money demand specifications were set to a fixed lead and lag of one. For the deterministic regressors, besides the constant, we also include a linear trend and a dummy variable. The presence of structural break is validated by Bai and Perron (1998) test for multiple breakpoints in both Eq.3 and Eq.4, and based on the test results, two breaks are detected, 2012Q1 and 2015Q2 respectively. The results are reported in Table A2 of the appendix.

The results presented in Table 2 shows that, as expected from the theory, the short-term interest rate represents the opportunity cost of holding money as the interest elasticity and interest semi elasticity of money demand are inversely related to the money income ratio. Both elasticities are also statistically significant at one percent level of significance. The adjusted R square for both money demand models shows that variation of the independent variables explains more than 80 percent variation in the log of money income ratio, with the adjusted R-squared of the double-log model slightly higher than semi-log model. In addition, the estimated models pass all diagnostic tests except the residuals that are serially autocorrelated. However, we obtain unbiased standard error estimates using Newey and West (1987) method.

Table 2: Nominal interest rate and money income ratio [18]

Variables	Log log model		Semi log model	
	Coefficients	P-value	Coefficients	P-Vakue
Intercept	-0.7976	[0.000]***	-0.9840	[0.0000] ***
Trend	0060	[0.0000]***	0.0060	[0.0000]***
Dum	0.0639	[0.0073] ***	0.0754	[0.0055]***
Lntbr	-0.2043	[0.0000] ***	-0.0296	[0.0002]***
$\sum_{i=-k}^{i=k} \vartheta_i \Delta X_{t+i}$			$\sum_{i=-k}^{i=k} \vartheta_i \Delta X_{t+i}$	
Goodness of fit				
Adjusted R-Squared:	0.8454		0.8405	
SSR	0.1430		0.1475	
AIC	-2.7895		-2.7580	
SIC	-2.5268		-2.4953	
Diagnostic tests				
Breusch Pagan LM test	Prob chi-sq(6)=0.3011		Prob chi-sq(6)=0.2214	
Breusch Godfrey LM test	Prob chi-sq(2)=0.0000		Prob chi-sq(2)=0.0000	
Jarque Bera test	Prob=0.4433		Prob chi-sq(2)= Prob=0.8095	
Ramsey reset test	Prob F (1,44)=0.9596		Prob F (1,44)=0.8547	

*** (**) [***]: Statistically significant at a 10 (5) [1] % level**

Moreover, we test whether the cointegrating regression DOLS is not spurious for both money demand specification using Phillips & Ouliaris (1990) and Engle & Granger (1987) cointegration tests. The two tests reject the null hypothesis of no cointegration for both specifications of money demand at 10 percent level of significance. Hence, we conclude that there is indeed a stable linear long-run linear relationship between money income ratio and nominal short-term interest rate for both money demand specifications.

[18] Probabilities value are reported in brackets

Table 3: ARDL Long run interest elasticities and Diagnostic tests [19]

Variables	Log-log model		Semi log model	
	Coefficients	P-value	Coefficients	P-Value
Intercept Trend Dum Intbr	-0.4937 0.0037 0.0269 -0.0792	[0.0003]*** [0.0002]*** [0.0987]* [0.0000]***	-0.5656 0.0037 0.0294 -0.0103	[0.0001]*** [0.0003]*** [0.0889]* [0.0220]**
Goodness of fit				
Adjusted R-Squared: SSR AIC SIC	0.8974 0.1045 -3.1641 -2.9411		0.8946 0.1075 -3.1364 -2.9134	
Diagnostic tests				
Breusch Pagan LM test Breusch Godfrey LM test Jarque Bera test Ramsey reset test	Prob chi-sq(6)=0.7056 Prob chi-sq(2)=0.7182 Prob chi-sq(2)=0.7668 Prob F (1,46)=0.6321		Prob chi-sq(6)=0.7337 Prob chi-sq(2)=0.8280 Prob chi-sq(2)=0.7436 Prob F (1,46)=0.7360	

*** (**) [***]: Statistically significant at a 10 (5) [1] % level**

Furthermore, to test the robustness of DOLS estimates, we also estimate the money demand elasticities using the alternative methodology, ARDL. The results are displayed in Table 3 above. First, Akaike information (AIC) criterion is used to obtain the optimal lag length for the ARDL model for both money demand specifications [20]. Then, a bound test for cointegration is used to test for the presence of the long-run relationship between money demand and interest rate for both money demand specifications. The F statistic of 10.47 for the double log model and 9.55 for the semi log model exceed the bound test critical value for I(1) variables at 5 percent level of significance[21]. The long-run interest elasticity obtained using ARDL is 0.0792 and 0.0103 for interest elasticity and semi-interest elasticity, respectively. Both elasticities are statistically significant at 1 percent level of significance and their negative sign clearly indicates a tradeoff between short-term interest rate and money income ratio. Note that the elasticities obtained using ARDL are lower than DOLS estimates and the goodness of fit improves significantly with the ARDL results.

[19] Probabilities value are reported in brackets
[20] AIC optimal lags is ARDL(2,0) for both Double log and Semi-log models.
[21] Bound test critical values at 5 percent level is 7.785 for n=55

We then proceed and estimate the welfare cost of inflation using the regression parameters obtained using DOLS and ARDL[22].

5.2. Welfare cost estimates

Assuming that a steady-state real interest rate (r) of 5.91 percent, which is the average real interest rate over the sample period (form second quarter of 2006 to the fourth quarter of 2019), represents the optimal rate at the inflation rate equal to zero and r=6.91% is equivalent to an inflation rate of 1 percent, the welfare cost estimates for different scenarios are presented in table 4 below:

Table 4: Welfare cost of inflation as percentage of GDP

Inflation rate	Nominal interest rate	Double Log model		Semi Log model	
		(DOLS) Parameters $A=\exp(-0.7976)$ $\eta=0.2043$	(ARDL) Parameters $A=\exp(-0.4937)$ $\eta=0.0792$	(DOLS) Parameters $B=\exp(-0.9840)$ $\xi=0.0296$	(ARDL) Parameters $B=\exp(-0.5656)$ $\xi=0.0103$
0	$r=0.0591$	1.2180	0.3882	0.0019	0.0010
1	$r=0.0691$	1.3794	0.4483	0.0026	0.0014
2	$r=0.0791$	1.5360	0.5077	0.0035	0.0018
3	$r=0.0891$	1.6887	0.5665	0.0044	0.0023
5	$r=0.1091$	1.9839	0.6826	0.0066	0.0035
8	$r=0.1391$	2.4070	0.8537	0.0107	0.0057
9	$r=0.1491$	2.5437	0.9101	0.0123	0.0065
10	$r=0.1591$	2.6785	0.9662	0.0140	0.0074

The results displayed in Table 4 show that the welfare cost of inflation of the double log model generates substantial welfare loss of GDP compared to the semi-log model for both DOLS and ARDL estimates. For instance, using different inflation rate scenarios that range from 2 percent to 8 percent, the welfare cost of inflation at one percent is equivalent to 1.5360 and 0.5077 percent loss in GDP using DOLS and ARDL estimates, respectively. At an inflation rate of 8%, the DOLS estimates generate the welfare cost of 2.4070 percent of GDP while the estimates of ARDL produce welfare cost estimates of 0.8537 percent of GDP. At the current target rate of 5% rate of inflation, the welfare cost of inflation is 1.9839 percent of GDP for DOLS and 0.6826 percent for ARDL.

[22] Note that the diagnostic tests for both specification of money demand indicated that the residuals are normal, serial uncorrelated and homoscedastic. Moreover, the Ramsey rest test fail to reject the null of no misspecification.

With the semi-log specification is associated with smaller welfare losses, with a slight increase in inflation rate from zero to 2 and 8 percent equalling 0.0035 percent of GDP and 0.0107 percent of GDP, respectively, for the DOLS procedure. In comparison to ARDL welfare cost estimates, the same rate of inflation produces 0.0018 percent loss in GDP for the former and 0.0057 percent loss in GDP for the latter. Using the current target rate of 5%, the DOLS semi-log model produces a welfare cost of 0.0066 percent loss in GDP while ARDL welfare loss is equivalent to 0.0035 percent of GDP.

This shows that if the correct money demand specification in Rwanda is double log model, the welfare cost of inflation is substantial compared to the semi-log model. For policy purposes on the optimal level of inflation, either one has to rely on the double log or the semi-log welfare estimates of DOLS or ARDL. For that reason, we perform in-sample fitting of both models. The results are portrayed in Table 5 below.

Table 5: Double log and Semi log Model performance

Accuracy measures	Double log		Semi-log	
	DOLS	ARDL	DOLS	ARDL
RMSE (Root mean Square Error)	0.0524	0.0444	0.0533	0.0450
MAE (Mean Absolute Error)	0.0396	0.0366	0.0406	0.0370
Theil U2 Coefficient	0.9470	0.8206	0.9708	0.8312

According to the results, the double log model predicts better the response variable (log of money income ratio) since the three measures have the lowest values compared to the semi log model for both DOLS and ARDL methods. We decided to rely more on welfare cost estimates of the double log model. Moreover, if one compares the predictability of ARDL and DOLS, the ARDL predicts better the log of money income ratio.

5.3. Discussion

The preliminary ARDL results of the double log specification suggest that, within the inflation benchmark band of 2 to 8 percent, the welfare cost of inflation ranges between 0.507 percent to 0.853 percent of real GDP per year. These estimates are slightly higher compared to those found in countries in the region with comparable economies such as Kenya. For example, a 5 percent inflation rate is associated with a 0.075 percent loss in real GDP every year in Kenya (Ikikii, 2017), which is much lower compared to the yearly loss of 0.6826 percent of real GDP in Rwanda.

A plausible explanation of the differential effect of inflation resides in the country-specific characteristics of money demand. As demonstrated by recent studies such as Mogliani and Urga (2018), Berentsen et al. (2015) and Cao et al. (2020), higher interest elasticity of money demand could reflect the historically lower level of uptake of financial innovation, which could be the case of Rwanda. This finding is in line with the results of Dunne and Kasekende (2018) that placed Kenya among African countries with higher growth rate of financial innovation and hence lower demand for cash.

6. CONCLUSION

Inspired by the global phenomenon of transition towards inflation targeting regimes among central banks, the National Bank of Rwanda recently adopted a price-based monetary policy framework. With price stability as the main objective of monetary policy, the central bank committed to maintaining a medium-term inflation benchmark of 5±3 percent using discretionary changes in its central bank rate. The benchmark was set in a bid to harmonize with key macroeconomic convergence criteria underlying the establishment of the EAMU.

The objective of this paper is to provide an estimate of the welfare cost of the anticipated inflation (5± 3 percent). Using quarterly observations from 2006:3 to 2019:4, two alternative money-demand function specifications are estimated, namely the double log and the semi-log models. The DOLS and ARDL procedures are used to obtain cointegrating regressions for both specifications. Our diagnostic tests suggest that, compared to the semi-log specification, the double log demand function performs better on the Rwanda data. Estimates of the double-log specification from the ARDL procedure turn out to have more predictive power and suggest a welfare cost ranging from 0.5077 to 0.8537 percent of annual real GDP over the medium-term inflation benchmark band. These estimates are substantially higher compared to what is found in the empirical literature, including case studies conducted in other countries in the region (such as Kenya).

These findings support the continued pursuance of price stability as the primary goal for monetary policy in Rwanda. Ensuring price stability would be the most efficient contribution of monetary policy to economic growth in Rwanda. The relatively higher interest rate elasticity of money demand and associated welfare losses imposed by inflation imply that enhanced policy strategies for financial innovation could be effective in reducing the burden of inflation on money holders, which support ongoing policy efforts of promoting innovative financial products to fast track financial access and usage of financial services in Rwanda. Failure to fast-track financial inclusion would support the case of rethinking the level and width of the current medium-term inflation objective of 5±3 percent, in favor of a lower and (perhaps narrower) benchmark.

Future empirical studies in Rwanda should address other and more indirect channels through which inflation affects the efficiency of resource allocation by exploring the potential of general equilibrium frameworks.

Moreover, there is a need for re-assessing the distribution of welfare costs in time and space. Given the recent developments in Rwanda's financial innovation, further empirical studies should provide alternative measures of welfare cost of inflation that account for instabilities in the long± run money demand. Researchers should also recognize that the welfare cost of inflation affects disproportionally certain categories of households, and thus depends on the distribution of socio-economic characteristics. Such insights can be unveiled by using micro-level data.

Last, it is worth mentioning that the findings of studies on the welfare cost of inflation would provide only a partial basis for judging the desirability of the inflation benchmark for Rwanda. There is a need to investigate other criteria of inflation benchmark of monetary policy such as business cycle fluctuations, wage and price rigidities, measurement error, monetary policy flexibility, zero interest rate bound, and the likes.

REFERENCES

Alvarez, F. & Lippi, F., 2009. Financial innovation and the transactions demand for cash. Econometrica, Volume 77, p. 363–402.

Arrau, P., De Gregorio, J., Reinhart, C. M. & Wickham, P., 1995. The demand for money in developing countries: Assessing the role of financial innovation. Journal of Development Economics, Volume 46, pp. 317-340.

Bai, J. & Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica, Volume 66, pp. 47-78.

Bailey, M. J., 1956. The welfare cost of inflationary finance. Journal of Political Economy, Volume 64, p. 93–110.

Bakhshi, H., Haldane, A. & Hatch, N., 1999. Some costs and benefits of price stability in the United Kingdom. In: M. Feldstein, ed. The Costs and Benefits of Price Stability. s.l.:University of Chicago Press, pp. 133 - 198.

Balls, E. & O'Donnell, G., 2002. The benefits of maintaining low and stable inflation. In: E. Balls & G. O'Donnell, eds. Reforming Britain's Economic and Financial Policy: Towards Greater Economic Stability. London: Palgrave Macmillan, pp. 58-70.

Barro, R. J., 2013. Inflation and economic growth. Annals of Economics and Finance, Volume 14, p. 85–109.

Berentsen, A., Huber, S. & Marchesiani, A., 2015. Financial innovations, money demand, and the welfare cost of inflation. Journal of Money, Credit and Banking, Volume 47, pp. 223-261.

Cagan, P., 1956. The monetary dynamics of hyperinflation. In: M. Friedman, ed. Studies in the Quantity Theory of Money. Chicago: University of Chicago Press, pp. 25-117.

Calza, A. & Zaghini, A., 2011. Welfare costs of inflation and the circulation of US currency abroad. The B.E. Journal of Macroeconomics, Volume 11, pp. 1-21.

Cao, S., Meh, C. A., Ríos-Rull, J.-V. & Terajima, Y., 2018. The Welfare cost of inflation revisited: The role of financial innovation and household heterogeneity. Bank of Canada Staff Working Paper, August.

Cao, S., Meh, C. A., Ríos-Rull, J.-V. & Terajima, Y., 2020. The Welfare cost of inflation revisited: The role of financial innovation and household heterogeneity. Journal of Monetary Economics, August.Volume In press.

Chadha, J., Haldane, A. G. & Janssen, N. G. J., 1998. Shoe-leather Costs Reconsidered. The Economic Journal, Volume 108, pp. 363-382.

Cysne, R. P. & Turchick, D., 2010. Welfare costs of inflation when interest-bearing deposits are disregarded: A calculation of the bias. Journal of Economic Dynamics and Control, Volume 34, pp. 1015-1030.

Dotsey, M. & Ireland, P., 1996. The welfare cost of inflation in general equilibrium. Journal of Monetary Economics, Volume 37, pp. 29-47.

Dotsey, M. & Ireland, P., 1996. The welfare cost of inflation in general equilibrium. Journal of Monetary Economics, Volume 37.

Driffill, J., Mizon, G. & Ulph, A., 1990. Costs of inflation. In: B. Friedman & F. Hahn, eds. Handbook of Monetary Economics. s.l.:Elsevier Science Publishers, pp. 725-1311.

Dunne, J. P. & Kasekende, E., 2018. Financial innovation and money demand: Evidence from Sub-Saharan Africa. South African Journal of Economics, Volume 86, pp. 428-448.

Elliott, G., Rothenberg, T. J. & Stock, J. H., 1996. Efficient tests for an autoregressive unit root. Econometrica, Volume 64, pp. 813-836.

Engle, R. F. & Granger, C. W., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica, Volume 55, pp. 251-276.

Feldstein, M., 1979. The welfare cost of permanent inflation and optimal short-run economic policy. Journal of Political Economy, 87(4), pp. 749-768.

Friedman, M., 1969. The optimum quantity of money. In: M. Friedman, ed. The Optimum Quantity of Money and Other Essays. Chicago: Aldine.

Gichondo, A., Karangwa, M., Kasai, N. & Munyankindi, P., 2018. Optimal inflation for economic growth in Rwanda. BNR Economic Review, Volume 13, pp. 27-43.

Gómez, V. & Maravall, A., 1996. Programs TRAMO and SEATS: Instructor for the user (Beta version: September 1996). Banco de España Documento de Trabajo.

Granger, C. & Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics, Volume 2, pp. 111-120.

Gupta, R. & Uwilingiye, J., 2008. Measuring the welfare cost of inflation in South Africa. South African Journal of Economics, Volume 76, pp. 16-25.

Gupta, R. & Uwilingiye, J., 2009. Measuring the welfare cost of inflation in South Africa: A reconsideration. South African Journal of Economic and Management Sciences, Volume 12, pp. 137-146.

Gupta, R. & Uwilingiye, J., 2010. Evaluating the welfare cost of inflation in a monetary endogenous growth general equilibrium model: The case of South Africa. International Business & Economics Research Journal, Volume 9, pp. 101-112.

Halicioglu, F., 2007. The financial development and economic growth nexus for Turkey. EERI Research Paper Series.

Hamilton, J. D., 1994. Time Series Analysis. Princeton: Princeton University Press.

Harris, R. & Sollis, R., 2003. Applied Time Series Modelling and Forecasting. 1st ed. Hoboken, New Jersey: Wiley.

Horvath, R. & Mateju, J., 2011. How are inflation targets set?. International Finance, Volume 14, p. 265–300.

Ikikii, S. M., 2017. The welfare cost of inflation in Kenya. Developing Country Studies, Volume 7, pp. 37-45.

Ikikii, S. M., 2017. The welfare cost of inflation in Kenya. Developing Country Studies, Volume 7, pp. 37-45.

Ireland, P. N., 2009. On the welfare cost of inflation and the recent behavior of money demand. The American Economic Review, Volume 99, pp. 1040-1052.

Ireland, P. N., 2009. On the welfare cost of inflation and the recent behaviour of money demand. American Economic Review, 99(3), p. 1040–1052.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, Volume 59, pp. 1551-1580.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford: Oxford University Press.

Keele, L. & Kelly, N. J., 2006. Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. Political Analysis, Volume 14, pp. 186-205.

Lamb, C., 1993. A rollercoaster out of control. Financial Times, 22 February.

Laurens, B. et al., 2015. The Journey to Inflation Targeting: Easier Said than Done, Washington D.C.: International Monetary Fund.

Leigh-Pemberton, R., 1992. The case for price stability. Bank of England Quarterly Bulletin, Volume 32, pp. 441-448.

Lucas, R. E., 2000. Inflation and welfare. Econometrica, Volume 68, p. 247–274.

Makochekanwa, A., 2008. Measuring the welfare cost of inflation in Zimbabwe. Munich Personal RePEc Archive Paper.

Makochekanwa, A., 2010. Measuring the welfare cost of inflation in Zimbabwe. Munich Personal RePEc Archive.

Marcellino, M., 1999. Some consequences of temporal aggregation in empirical analysis. Journal of Business & Economic Statistics, Volume 17, pp. 129-136.

Marty, A. L., 1999. The welfare cost of inflation: A critique of Bailey and Lucas. Federal Reserve of St Louis Review, pp. 41-46.

Meltzer, A. H., 1963. The demand for money: The evidence from the time series. Journal of Political Economy, Volume 71, pp. 219-246.

Mogliani, M. & Urga, G., 2018. On the instability of long-run money demand and the welfare cost of inflation in the United States. Journal of Money, Credit and Banking, Volume 50, pp. 1645-1660.

Mushtaq, S., Rashid, A. & Qayyum, A., 2012. On the welfare cost of inflation: The case of Pakistan. The Pakistan Development Review, Volume 51, pp. 61-96.

Newey, W. & West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, Volume 55, pp. 703-708.

Pakko, M. R., 1998. Shoe-leather costs of inflation and policy credibility. Review, pp. 37-50.

Panopoulou, E. & Pittis, N., 2004. A comparison of autoregressive distributed lag and dynamic OLS cointegration estimators in the case of a serially correlated cointegration error. Econometrics Journal, Volume 7, pp. 585-617.

Pesaran, H., Shin, Y. & Smith, R. J., 2001. Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, Volume 16, pp. 289-326.

Pesaran, M. H., Shin, Y. & Smith, . R. J., 2001. Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, Volume 16, pp. 289-326.

Phillips, P. C. & Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. Econometrica, Volume 58, pp. 165-193.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory, Volume 7, pp. 1-21.

Serletis, A. & Yavari, K., 2004. The welfare cost of inflation in Canada and the United States. Economics Letters, Volume 84, p. 199–204.

Serletis, A. & Yavari, K., 2005. The welfare cost of inflation in Italy. Applied Economics Letters, Volume 12, p. 165–168.

Serletis, A. & Yavari, K., 2007. On the welfare cost of inflation in Europe. Applied Economics Letters, Volume 14, p. 111–113.

Shah, I. A., Agarwal, M. L. & Kundu, S., 2019. Welfare cost of inflation: Evidence from India. Journal of Quantitative Economics, Volume 17, p. 781–799.

Stock, J. H. & Watson, M. W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica, Volume 61, pp. 783-820.

Tuan, N. A. & Luan, N. T., 2017. Money demand and the welfare cost of inflation: Empirical evidence from East Asian countries. Journal of International Economics and Managment, Volume 94, pp. 1-8.

Watanabe, T. & Yabu, T., 2018. The demand for money at the zero interest rate bound. CARF Working Paper.

Yoshino, J. A., 1993. Money and banking regulation: The welfare costs of inflation, Chigago, IL: s.n.

APPENDIX

Table A1: Unit root tests

Series	Model	DF_GLS(1996)	KPSS(1992)	Conclusion
Lntbr	τ_τ τ_μ τ	-3.25*** -2.43***	0.05*** 0.30***	Stationary
D(lntbr)	τ_τ τ_μ τ			
tbr	τ_τ τ_μ τ	-3.43*** -2.53***	0.05*** 0.28***	Stationary
Lnrm1_sa	τ_τ τ_μ τ	-2.86*** -0.35***	0.07*** 0.35	Non-stationary
D(lnrm1_sa)	τ_τ τ_μ	-5.54*** -5.49*** -	0.05*** 0.06***	Stationary

Notes: τ_τ represents the trend and intercept, τ_μ the intercept, τ no constant no trend and *(**)[***] indicate significance at 1%, 5% and 10% level of significance, respectively.

Table A2: Unit root tests break points

Series	Model	DF_GLS(1996)	Breaks	Conclusion
Lntbr	τ_τ τ_μ	-3.55 -3.55	2014q4 2013q2	No-Stationary
D(lntbr)	τ_τ τ_μ	-6.28*** -6.32***		
Tbr	τ_τ τ_μ	-4.34 -4.34	2013q2 2014q1	Stationary
D(tbr)	τ_τ τ_μ	-5.14*** -5.14***		Non-stationary
lnrm1_sa)	τ_τ τ_μ	-3.55*** -2.50***	2016q4 2013q1	Non-Stationary
D(lnrm1_sa)	τ_τ τ_μ	-6.27*** -6.32***		Stationary

Notes: τ_τ represents the trend and intercept, τ_μ the intercept, τ no constant no trend and *(**)[***] indicate significance at 1%, 5% and 10% level of significance, respectively.

Table A3: Multiple Break Tests Bai and Perron (Sequentially determined breaks)

Break Test	F-statistic	Scaled F-statistic	Critical Value**
0 vs. 1 *	138.3340	138.3340	8.58
1 vs. 2 *	23.54897	23.54897	10.13
2 vs. 3	3.676071	3.676071	11.14
Break dates	Sequential	Repartition	
1	2012Q1	2012Q1	
2	2015Q2	2015Q2	

* Significant at the 0.05 level. ** Bai-Perron (Econometric Journal, 2003) critical values.

SHORT-TERM INFLATION FORECASTING MODELS FOR THE NATIONAL BANK OF RWANDA

Faustin Maniraguha[23]
Christian Manishimwe[24]
Mathias Karangwa[26]
Patrick Hitayezu[27]

[23] Senior Economist, Research Department, National Bank of Rwanda
[24] Senior Economist, Monetary Policy Department, National Bank of Rwanda
[25] Senior Principal Economist, Research Department, National Bank of Rwanda
[26] Director, Research Department, National Bank of Rwanda

ABSTRACT

The paper documents the tools used by the National Bank of Rwanda for short term inflation forecasting and provides a set of information related to short-term inflation forecast and its combination to get accurate and one single forecast to inform monetary policy authorities. The paper employs benchmark models (ARMA and Random walk) and multivariate models (VARs and BVARs) to estimate headline inflation throughout a disaggregation process on its main components that are core, food and energy. The findings suggest that multivariate models outperform the benchmark models in terms of forecast accuracy and they indicate that the forecasting combination reduces forecast error compared to the individual model forecasts. Therefore, in line with monetary policy orientation under a forward-looking framework, the paper recommends the continuous use of combined multivariate models for short-term inflation forecasts as well as to consider their forecasts in feeding the quarterly projections model (QPM) as they include more information on economic dynamics than statistical models.

Key Words: Inflation, Short-term Forecasting Models, Forecast combination
JEL Classification: C52, C53, E31, E37

1. INTRODUCTION

The main mission of the National Bank of Rwanda (NBR) is to ensure price stability and a sound financial system. For any central bank, maintaining price stability is a clear but daunting task due to the fact that monetary policy decisions do not affect the real economy (i.e. inflation and output) instantaneously. Given these lags in the monetary policy transmission mechanism, the general consensus is that monetary policy should take forward-looking and medium-term perspectives (Lars, 2000) The common practice is to set a medium-term inflation objective and to implement the required short-term policy interventions to attain this objective. The medium-term inflation objective is set basing on model-based macroeconomic projections, which guide the deliberations of the monetary policy decision-makers. In addition to mitigating the effects of macroeconomic shocks, short-term policy interventions are also based on model-based macroeconomic projections and serve the purpose of re-aligning the path of macroeconomic variables to the desired path to ensure that the medium-term inflation objective is attained (Lars, 2000).

The NBR previously had a point medium-term inflation objective of 5%, which was later revised to a target band of 5%±3%. The medium-term inflation objectives are set on the basis of model-based evidence on the current and projected state of the economy, which guides the discussions of monetary policy decision-makers. Each quarter and whenever deemed necessary, the MPC sits to review macroeconomic developments and identify if there are upward or downward risks to the inflation outlook and takes appropriate decisions if needed.

The decisions of the MPC are then communicated to the markets and this helps to anchor inflation expectations of firms and households, thereby making the central bank more effective in fulfilling its objective. Consequently, forecasts for inflation, output and other macroeconomic variables are an essential input in the monetary policy decision-making process.

In order to be more effective in fulfilling its mandate, the NBR has since 2009 been building and strengthening analytical capacity to support evidence-based forward-looking monetary policy decision-making by developing the appropriate modeling and forecasting tools and conducting expectations surveys. Since the creation of the modeling and forecasting function in 2009, both near-term or short-term and medium-term forecasting models have been created and used. As time goes, such models continue to be improved in line with macroeconomic dynamics.

Furthermore, the NBR adopted a "Forecasting and Policy Analysis Systems (FPAS)", a system designed to help policymakers in decision-making and communicating decisions, through building capacity for macro-economic analysis and forecasting, integrating forecasting and empirical analysis into decision making and improving internal and external communication. This has led to the introduction of the forecasting team, the forecasting round and the restructuring of the Monetary Policy Committee (MPC) membership to include an external member. The typical forecasting round at NBR lasts around six weeks and includes an MPC meeting that is preceded with 2nd pre-MPC meetings which take the form of structured discussions between MPC members and the forecasting team on economic developments as well as near and medium-term outlook. The MPC decisions are implemented by the Monetary Policy Implementation Committee (MPIC) as well as by the Financial Markets Operations Committee (FMOC).

Forecasting at NBR is a core component of the FPAS and inflation is forecasted over two-type horizons: the short-term horizon of one to six months, where both univariate and multivariate models are used, and, the medium-term horizon, where the quarterly projections model (QPM) is the main forecasting tool. The QPM is a small-scale New-Keynesian macroeconomic model, which not only serves for medium-term projections but also guides policy discussions and alternative policy simulations. The role of near-term forecasting has been emphasized with the introduction of the QPM model.

Consequently, the production of near-term inflation forecasts serves for data collection and data processing, prediction of exogenous variables, analysis of actual economic development and initial conditions, comparative benchmark up to 3 quarters and the prediction of variables that are not included in the core model. Given this tremendous role of near-term/short-term inflation forecasts, it is very important that the outcome is accurate. In this respect, accurate forecasts of future inflation over the medium term depend on precise information from current inflation, as well as precise, non-biased forecasts resulting from short-term inflation models.

Therefore, since decisions making is contingent on the quality of forecasts, it is in the interest of central banks to evaluate and improve the performance of their forecasting models.In the early stages, the NBR relied on univariate models, such as ARIMA (Autoregressive Integrated Moving Average) and STIF (Short-term Inflation Forecasting) for short-term inflation forecasting (Mwenese & Kwizera, 2018). These models had certain drawbacks, such as being atheoretical (i.e. not based on economic theory and thus unable to address issues such as endogeneity) and having strong memory of the past. In view of this, some multivariate models were introduced. However, the multivariate models also did not consider external inflation factors, such as international food and oil prices. By construction, a set of models produced different forecasts, calling for the development of new methods to evaluate and combine forecasts from these different forecasting models.

Unlike previous studies on inflation in Rwanda (Gichondo & Kimenyi, 2012; Mwenese & Kwizera, 2018; Maniraguha et al., 2019; Ruzima & Veerachamy, 2015; Nyoni & Solomon, 2018; Kelikume & Salami, 2014; Gupta & Kabundi, 2008; Doguwa & Alade, 2013; Krušec, 2007), this paper focuses on estimating and forecasting each component of total headline inflation that includes core, food and energy. For inflation modeling purposes, the estimation refers to core excluding food inflation having 61.3 weights in total headline CPI, food inflation that has 32 weights and energy inflation that contributes 6.7 percent in total headline CPI. According to (Rummel, 2015), a disaggregated approach expands the information set and in doing so, it leads to better forecast performance in the short term as it takes into account sector-specific information and helps to understand sectoral inflationary pressures as well as current price dynamics.In the spirit of improving the short-term inflation forecasts accuracy, the paper discusses short-term inflation forecasting models using different modeling approaches, and how to evaluate and combine the forecasts to ensure consistency and easy communication to policymakers. Thus, the contribution of the paper is twofold: (1) to document the tools used by at the National Bank of Rwanda for short-term inflation forecasting, (2) to provide a set of information vis-à-vis short term/near term inflation forecast and its combination to get accurate forecast to inform monetary policy authorities.

Besides this introduction, the remainder of this paper is divided into five (5) sections: section two contains a brief literature review with a focus on the short-term inflation estimation and forecasting models, forecast performance evaluation and combination. Section three describes the methodology, variables selection and data, section four discusses the empirical findings, while section five gives the conclusion, along with the summary and policy implications.

2. LITERATURE REVIEW

This section sheds light on both theoretical and empirical views underlying the short-term inflation estimation and forecasting models as well as forecast combination. The section is not exhaustive to explain the short-term inflation models but models like Random walk, Autoregressive and Moving Average (ARMA), Vector Autoregressive (VAR) and Bayesian Vector Autoregressive (BVAR) models are covered in both theoretical and empirical folds. Moreover, forecast combinations is part of the literature review for a better understanding.

2.1. Theoretical literature review

Theories of inflation started over the decades with different economic thoughts. The first is the monetarist theory represented by the Chicago school or Monetarist school led by Milton Friedman. This theory states that the general increase in price level is always and everywhere a monetary phenomenon (Nyoni & Solomon, 2018). This means that inflation results from the variations in the quantity of money within the economy.

Other economic thoughts related to the drivers of inflation are centered on the Neo Keynesian school views and inflation expectations. Starting from Friedman & Bordo, (2006), and Addison and Burton (1980), demand-pull inflation drives inflation. This means that an excess demand over supply creates pressures on inflation as a result of government, private sector and consumer spending (Doguwa & Alade, 2013; Nyoni & Solomon, 2018). According to Humphrey (1977) and Kavila & Roux (2017), inflation pressures may be created by high costs on the factors of production. This refers to the cost-push theory which describes that a surge in the inputs and production costs may lead to an increase in general price levels, i.e inflation (Nyoni & Solomon, 2018). In line with the drivers of inflation and differently from the previous theories, Bernanke (2005), and Khan and Schimmelpfennig (2006) indicated that inflation can result from structural rigidities in the economy where these can comprise land tenure, lack of storage facilities, poor harvest, and overdependence on rainfall.

Unlike the previous inflation theories, Muth (1961) brought the idea of rational expectation, which was more elaborated by Robert and Sargent (1979). According to this view, in the past time, economic agents used backward decision-making (adaptive expectation) while under rational expectation, economic agents may use all available information and conceive the trajectory of inflation. Such revolution in inflation expectation is integrated into macroeconomic policy using the augmented Philips curve relationship.

Since the theories on the causes of inflation differ from one school to another, monetary policy authorities always seek to achieve price stability by putting in place procedures and tools to model and predict inflation depending on each country's context.

The forward-looking aspect adopted by some Central Banks, of which National Bank of Rwanda, underpins the importance of forecasting inflation and other macroeconomic variables. Modeling and forecasting inflation involves the use of econometric models that are divided into two strands. The first strand includes statistical forecasting or univariate or extrapolation models. According to Jenkins et al. (2016), such models are backward-looking as they use past information to predict the future. In addition, they are easy to build and they are able to produce accurate projections for short-term horizons, despite being atheoretical (Spencer, 1993). The second strand comprises multivariate models. Unlike univariate models, these models are theoretically based and consist of equations or a system of equations with more than one variable. Lucas (1976), recommend using models based on theories in forecasting rather than looking at the statistical correlations analysis. To him, having a model-based theory is essential to policy formulation.

The literature on inflation forecasting uses various econometric techniques. Most studies extensively use univariate and multivariate techniques, such as ARIMA (Jenkins, et al., 2016), vector autoregressions (Bañbura, et al., 2010) and dynamic factor models (Stock & Watson, 2002). The latter are characterized by capturing rich dynamics among the variables. The ARIMA model is a tool for modeling and forecasting time series data based on their past values. A significant advantage of univariate ARIMA approach is that it can be developed in a relatively short time with automated State Space Forecasting algorithms, although it has limited explanatory capability. The vector autoregression (VAR) models are popular tools for forecasting and policy analysis among the possible multivariate time series models. These models allow the interaction of different related macroeconomic variables. They do not suffer from an endogeneity problem but they may suffer from the over parameterization problem as the number of parameters to be estimated increases geometrically with the number of variables and proportionally with the number of lags included, resulting in inaccurate estimation of parameters. The BVAR approach limits the dimensionality problem by shrinking the parameters into hyper-parameters via the imposition of priors, which makes it possible to estimate VAR models accurately.

2.1.1 Forecast performance evaluation

Once the forecasts from different models are produced, they need to be evaluated by testing their performance (Diebold & Lopez, 1996). Forecasting models are compared whereas their forecasts are judged basing on bias, precision and accuracy criterion. Therefore, a good forecast has to be unbiased, and should have high accuracy and high precision. There is a set of literature on the forecast performance evaluation and the techniques that are put in place to do so. These techniques include the mean error to judge the bias among forecasting models, Mean Square Error and error variance to test the precision.

The forecast accuracy is evaluated using different measures and they include the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE), Theil Inequality Coefficients (Theil U1 and U2). The literature on the techniques used to evaluate the forecasting models' accuracy include (Chai & Draxler, 2014); (Willmott, et al., 2009).

Though each technique has its own property, multiple techniques are needed to better confirm a good forecasting model at certain horizons.

2.1.2. Forecast combination

Usually, multiple forecasts are available to decision-makers and the differences in forecasts reflect differences in modeling approaches, multiple data sources, and differences in subjective priors. Therefore, policymakers are concerned with what forecast to pick between a single forecast or using some forecast combinations.

The literature led by Bates and Granger (1969), Granger and Ramanathan, (1984), Diebold and Mariano (2002), Stock and Watson (2002), Kapetanios et al. (2007), Bjørnland et al. (2008), Öðünç et al. (2013), Montero-Manso et al. (2018) and Atiya (2020) demonstrated how forecast combinations give better outputs in terms of accuracy compared to single forecasts. The forecast combination may be useful to deal with external forecasts (underlying model unobserved to policymakers), model uncertainty/misspecification, structural breaks and some other non-stationarities and large number of explanatory variables.

Empirically, different approaches were put in place concerning the forecast combination. They include averaging using a scheme where the forecast models are assigned equal weights and, forecast combination with weighted average. Referring to the work done by Timmermann (2006), and Smith and Wallis (2009), the scheme of equal weights is considered as optimal and hard to beat.

In the same view, Clark and McCracken(2010) argued that when forecasts are combined, the results become more robust to the aforementioned instabilities. Furthermore, the combination of forecasts allows modeling and forecasting of the misspecified models where some variables may be omitted, hence the combination may include all these assumed biases. In general cases, forecast combination is not supposed to be superior but it is preferable against choosing a single bad model (Bjørnland, et al., 2008).

The literature review on the short-term inflation forecasting models starts from the assessment on the source or drivers of inflation, models for estimation and forecasting coupled with forecasts evaluation and combinations. Multiple studies addressed the gap that linked to the variable selection in the model, understanding the best model appropriate for modeling and forecasting inflation, the procedures or steps to pass through so that to get accurate inflation forecasts.

Such studies used different types of data for instance quarterly and monthly data. Both statistical and multivariate models were applied depending on the country's context.
Here below is the table summarizing the empirical literature review concerning the short-term inflation estimation, forecasting, evaluation, adding to the forecasts combination exercises within national economies.

2.2. A summary of the empirical literature review on short-term inflation forecasting, evaluation and combinations.

Authors	Case study	Methods, variables, data, and sample size	Findings
(Gichondo & Kimenyi, 2012)	Rwanda	VARX Model was applied to study inflation process in Rwanda. The model used quarterly observations on CPI, GDP, exchange rate, and international oil price. Data spanned from 1997-2009.	Findings through the co-integration analysis indicated that economic growth works to reduce inflation while money supply (M3), exchange rate depreciation, international oil price affect inflation positively. In the short-run, lagged inflation proves to be the only significant variable explaining inflation. This confirms the relevancy of ARMA model that was in application in BNR as short-term inflation forecasting.
(Mwenese & Kwizera, 2018)	Rwanda	This study uses autoregressive models (ARMA and STIF: Short-term inflation-forecasting model disaggregating the 12 CPI headline components). In addition, a new Keynesian forecasting model, FPAS was used and helped to estimate the following equations: the IS curve[1], the Phillips curve, the UIP condition[2] and the monetary policy reaction function. The sample size comprises the comparison of the two forecasting rounds: 2017Q2 and 2018Q1.	The backward looking models: ARIMA, STIF and State-Space models, were in use for short-term forecasting, while a small-scale New-Keynesian macroeconomic, Forecasting and Policy Analysis System (FPAS) is used for the long horizon. The macro model was found powerful in the medium term and it helps to predict the trajectory of the all economy. Additionally, the ability of non-structural models decreases in presence of shocks.
(Maniraguha, et al., 2019)	Rwanda	GARCH family models were used to estimate time-varying volatility in consumer price, namely headline, food and non-alcoholic beverages, housing and transport CPI in Rwanda. This was applied on quarterly data spanning from 2007-2018.	The findings explained in details the magnitude and persistence in the volatility of the CPI inflation components. The study recommended the use of the results in enhancing the accuracy of the short-term inflation forecasts that are important inputs in the FPAS (Forecasting and Policy Analysis System), a quarterly projection model at the National Bank of Rwanda.
(Nyoni & Solomon, 2018)	Nigeria	Headline inflation in Nigeria was modeled and forecasted using ARMA, ARIMA and GARCH models.To do this, annual time series data on inflation rates in Nigeria from 1960 to 2016 was used.	Using Theil's approach in evaluating forecasts, the study showed the ARMA (1, 0, 2) model, the ARIMA (1, 1, 1) model and the AR (3) – GARCH (1, 1).Therefore, the study confirmed that ARMA (1, 0, 2) model was the optimal model among of them.
(Kelikume & Salami, 2014)	Nigeria	VAR and ARIMA were used to estimate and predict inflation for Nigeria. Month data spanning from 2003:01 to 2012:06 were used in this study.	The findings showed that VAR model specification outperformed the ARIMA model, taking into consideration the minimum root square errors.

[27] IS curve: Investment Saving Curve

[28] UIP condition: Uncovered Interest rate parity condition

(Gupta & Kabundi, 2008)	South Africa	The Dynamic Factor Model (DFM) framework was used to project per capita growth rate, inflation, and the nominal short-term interest rate for the South African economy. The model uses 267 quarterly observations starting from 1980Q1 to 2006Q4.	Basing on the RMSE as an indicator to evaluate forecasts, the results showed that DFM predicts per capita growth rate, nominal short-term interest rate and inflation better than the NKDSGE.
(Doguwa & Alade, 2013)	Nigeria	Around four short-term headline inflation-forecasting models were used and they were based on the SARIMA and SARIMAX. Models were applied to headline, food and core CPI equations. Data spans from July 2001 to September 2013.	The findings revealed that SARIMA is appropriate to produce forecasts of food inflation for less than 10months ahead while SARIMAX is powerful for eleven and twelve months ahead. In addition, SARIMA is better than SARIMAX in producing the forecasts of core inflation.
(Krušec, 2007)	Slovenia	The constructed short term models included AR process, VAR and factor models on the overall inflation and the subcomponents (energy inflation, Industrial goods inflation, services inflation, processed food and the non-processed food inflation). Y-o-Y inflation rate data were considered and they spanned from 1997 to year 2001.	Using the RMSE to judge the best models, the findings indicated that factor models performed better than AR benchmark model and somehow similar to the VAR models on all inflation components.
(Gupta, et al., 2012)	South Africa	The study used both VAR and BVAR to forecast macroeconomic variables including consumer price index inflation rate. The study employed data spanning from 1960Q1 to 199Q4 while out of sample covered 200Q1-2011Q2.	Throughout the forecast performance evaluation, both models have equal capability. Thus, the study recommended extending variables in the model from 3 variables to 10 variables.
(Kapetanios, et al., 2007)	Bank of England	Linear3 and nonlinear univariate models4, vector autoregressive (VAR) models of various specifications5, Bayesian VARs (BVARs), factor models and time-varying coefficient models.	The forecast performance of GDP and inflation improved when the models are combined. The combined forecasts incorporate information from the entire range of models and data. To some extent, combined are robust for model misspecification.
(Bjørnland, et al., 2008)	Norges Bank	The inflation forecast combination included the following: autoregressive integrated moving average (ARIMA) models, a random walk (RW) in mean model, VAR models, Bayesian estimated VAR models, error correction models, monthly and quarterly factor models, and then, a dynamic stochastic general equilibrium (DSGE) Model.	The results revealed that one model may perform better in one horizon and become bad in another horizon. Similarly, the study demonstrated that there is an advantage of averaging forecasts from several individual models when predicting inflation in Norway in the short term (up to a year).

[29] Linear models includes: Unconditional mean, Random walk

[30] Non-linear models includes: Markov-switching model, Smooth-transition autoregressive model,

[31] They include: Autoregressive, vector autoregressive, Vector autoregressive model, monetary, Double differenced vector autoregressive model, Double-differenced vector autoregressive model, monetary, Recursively estimated vector autoregressive model: small data set, Recursively estimated vector autoregressive model: large data set

(Öðünç, et al., 2013)	Turkey	Some short-term inflation models were combined and they included both univariate and multivariate models that are: Unconditional mean, Random walk, Seasonal ARIMA, non-linear models (decomposition based models),a Phillips curve motivated time varying parameter model, a suite of VAR and Bayesian VAR models and dynamic factor models.	The findings showed that multivariate models as they have more economic information performed better than the benchmark models like random walk while the relative performance in producing good forecasts falls around 30% on average, mainly conspiring the first two quarters ahead. From the forecasting combinations, the findings demonstrated that forecast combination reduces forecast errors in comparison with the forecasts from single models though sometimes they can be better in some periods.
(Hsiao & Wan, 2014)	United States of America	In this study, some approaches in forecast combination include eigenvector approach, a mean corrected and trimmed eigenvector. These are used to minimize the mean squared prediction error when there is no structural change. Monthly data starting from 1960M3 to 2008M12 were analyzed.	The results showed that Bayesian averaging methods that combine the weights produces more accurate forecasts than the eigenvector method and the regression technique. In addition, the mean corrected simple average was found to produce forecasts that are properly robust in terms of mean squared forecast errors.
(Lack, 2006)	Swiss National Bank	The study used the VAR models that are Unrestricted VAR and Bayesian VAR to predict Swiss consumer price inflation. Data spanned from 1986Q4 to 2005Q, while RMSE and Theil U were applied to judge the forecast performance.	The results indicated that by combining forecasts, the quality of forecasts may be improved substantially.

In a nutshell, a few studies have concentrated on forecasting inflation using the disaggregated approach. This study contributes to the empirical literature by presenting short-term inflation forecasting through the disaggregation of headline consumer price indices into its main components: food, energy and core components. According to Rummel (2015), a disaggregated approach provide more accurate forecast than the aggregated one. The principle for this approach is to isolate the most heterogeneous or volatile series, that apparently are driven by specific explanatory factors. In respect of inflation, it is to separate the most volatile components such as food and energy from core inflation, then model and forecast each of these components separately. Thereafter, an aggregated forecast can be derived by combining the forecasts of these CPI components.

3. METHODOLOGY, VARIABLES SELECTION AND DATA

This study relies exclusively on the use of quantitative method, we describe various models used in inflation forecast, their performance evaluation and forecast combinations, thereafter, we go through the data, to shed light on the statistical/univariate and multivariate models used to forecast headline inflation and its main components (core, food and energy) at the National Bank of Rwanda. This section describes each model specification whether it is based on any theory or not. The section deals with forecast performance evaluation, to judge multivariate models against benchmarks models. Other issues covered under this section are forecast combinations, the choice of variables, and definition and explanation of the data used in this study.

For a proper understanding and clarification, some similar works are quoted in this section. Building on that, the Eviews codes for each model are given right after each model specification. Since headline CPI inflation is made up of three main components: food, core and energy, it is highly advantageous to model each component and then after use their respective weights to derive an aggregate headline CPI inflation forecast. This exercise is one way of combining forecasts.

Each model, be it statistical or multivariate specifies food, core and energy components. Important to note is that modeling the specific components of headline inflation gives an insight on the behavior of each component and the magnitude it may have in influencing headline inflation.

3.1. Model specification, forecast performance evaluation and forecast combinations

3.1.1. Model specification

a. Benchmark models

The benchmark models: Random Work (RW) and ARMA are the so-called statistical or univariate models. Though it is difficult to beat them in the very short-term, they are not considered in deriving the forecast average (i.e. forecast combination). They are only used to judge the performance of the multivariate models (Harvey, 1989). In addition, vis-à-vis the multivariate models, statistical models are beaten for the extended short-term periods, following their weak robustness in terms of root mean square errors.

a. 1 Random Walk (RW)

The (Pure) Random Walk (RW) model is one of the statistical models and is often called no change model. Such models are robust to common forms of structural change like intercept shifts. The model is specified as:

$$y_t = y_{t-1} + \varepsilon_t \tag{1}$$

The corresponding h-step ahead forecast is defined as follows:

$$E(y_{t+h\backslash t}) = y_t \tag{2}$$

Where: $E(y_{t+h\backslash t}) = E(y_{t+h\backslash} y_t, y_{t-1}\dots\dots\dots\dots\dots\dots)$ is the h-step ahead forecast.

Therefore, in line with the National Bank of Rwanda, a random walk model with drift for the three components of headline inflation starting from core, energy and food is modeled separately as follows:

$Y_t=c+Y_{t-1}+\varepsilon_t$; where Y_t is core, energy and food inflation respectively in the respective equation. C is a constant while ε_t is the white noise or random fluctuation at that time.

a. 2 Autoregressive Moving Average (ARMA): AR (p, q)

The autoregressive model is another type of statistical/univariate model. This model contains the order of lags and moving average of the shocks. The model is not robust to structural change but it is robust to misspecification resulting from the variable selection. This refers to the extrapolation approach and it is used to produce inflation forecasts for short-term periods, not more than six months ahead.

According to (Jenkins, et al., 2016), univariate models like ARMA possess certain properties: (1) detecting stationarity, (2) detecting seasonality, (3) differencing to achieve stationarity, (4) identification of the number of lags: AR (p), (5) as well as the identification of the moving average order: MA (q). To perform the ARMA process, firstly, the order of lags is specified and the moving average order follows. Here below is the order of lags: AR (p) specification:

$$x_t = c + \sum_{i-1}^{p} \alpha_i x_{t-i} + \varepsilon_t \tag{3}$$

Where x_t is the variable of interest. The lag p is picked following the information provided by econometric methods used to select lags. The information criteria are often used to guide model selection (Grasa, 1990). And these are Akaike Information Criterion (AIC), Schwarz Criterion (SIC), and the Hannan-Quinn Criterion (HQ). The variable of interest is explained by its own lags plus shocks at time t.

Regarding order of MA (q) specification, the equation become as follows:

$$x_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{4}$$

Where x_t is the variable of interest which is the level of moving average of the error terms or shocks and it depends on its previous levels. The order is determined by the selection criteria as mentioned above. In order to make the difference from (Mwenese & Kwizera, 2018) who directly estimated an aggregated headline consumer price index equation, here the component is disaggregated and estimated starting from core, energy and food components.

Empirically, the National Bank of Rwanda estimates the ARMA equations for core, energy and food inflation as written here below.

The process start with the determination of the order of both lags and moving average for each component and then an equation that provides a one-step-ahead forecast is estimated. The number of lags and moving average are subject to change as new data are released. Hence, the exercise of identifying the order of lags and moving average is done each forecasting round and whenever new data come in.

a. 2.1 Core inflation

$$\boldsymbol{\pi_t^{Core} = \varepsilon_t + \sum_{i=1}^{4} \theta_i \varepsilon_{t-i}} \tag{5}$$

From the above equations, core inflation is estimated with zero order of lags and moving average that starts from 1 to 4 orders. The number of orders are determined following the lag selection criteria among Akaike Information Criterion (AIC), Schwarz Criterion (SIC or BIC), and the Hannan-Quinn Criterion (HQ).

a.2.2 Energy inflation

Energy inflation is modeled with order of two lags and moving average. Both orders are chosen using the information selection criteria as in core inflation.

$$\boldsymbol{\pi_t^{Energy} = \pi_{t-1}^{Energy} + \pi_{t-2}^{Energy} + \varepsilon_t + \sum_{i=1}^{2} \theta_i \varepsilon_{t-i}} \tag{6}$$

a.2.3 Food inflation

Like core and energy inflation, food inflation estimation includes the order of three lags while moving average contains two orders. Both orders are chosen using the information selection criteria as in core and energy inflation.

$$\boldsymbol{\pi_t^{Food} = \pi_{t-1}^{Food} + \pi_{t-2}^{Food} + \pi_{t-3}^{Food} + \varepsilon_t + \sum_{i=1}^{2} \theta_i \varepsilon_{t-i}} \tag{7}$$

b. Multivariate models

In the construction of the multivariate models for the National Bank of Rwanda, some challenges were to be resolved: (a) data management as all variables used to produce inflation forecasts are loaded and imported from one single source called intermediate file, (b) inclusion of the endogeneity in the model specifications whereby Vector Autoregressive (VAR) and Bayesian

Vector Autoregressive (BVAR) are constructed, (c) consideration of exogenous factors whereby VARX and BVARX were also added.

Underneath this section, a set of multivariate models are covered and they include Vector Autoregressive (VAR), Vector Autoregressive with external variable (VARX), Bayesian vector Autoregressive (BVAR), Bayesian Vector Autoregressive with external variable (BVARX), Joint Vector Autoregressive (JVAR), Joint Vector Autoregressive with external variable (JVARX), Joint Bayesian Vector Autoregressive (JBVAR), Joint Bayesian Vector Autoregressive with external variable (JBVARX), Joint Bayesian Vector Autoregressive with relative prices (JBVAR_RP), Joint Bayesian Vector Autoregressive with external variable and relative prices (JBVARX_RP).

These models are used to estimate each component of headline inflation that includes core, food and energy components. Thus, to have a good forecast of headline inflation, the joint models are also produced per each model. This simply means that for instance a joint VAR model at the end includes three components, which are core, food and energy at the same time. Having a joint VAR, VARX, BVAR, BVARX and Joint BVAR and BVARX models with relative prices, it helps to come with a final forecast for headline core, food, energy and headline inflation.

Overall, VARs and other related models are amongst the accurate forecasting model classes available. This is due to the over-parameterization. In the work done by (Sims, 1980), multivariate models such as VAR have to be used for three (3) purposes: (a) forecasting economic time series, (b) designing and evaluating economic models, (c) evaluating the consequences of alternative policy actions.

The National Bank of Rwanda chose these types of models in the spirit of having strong models capable to forecast inflation in a period of two quarters ahead rather than relying on benchmark models. Still, in line with the current forward-looking price-based monetary policy framework, multivariate models are powerful to feed the quarterly projections models (QPM). Theoretically, inflation is influenced by its past levels but also some other domestic and external variables. Thus, all variables need to be modeled together with inflation. Any change in such variables, affect inflation either directly or indirectly. The variables coming in the model respect the properties of stationarity which is why they are differenced via a created Eviews program.

b.1 Simple Vector Autoregressive (VAR) model

A Vector Autoregressive model generally shows the dynamic relationship between variables. Let y_t be a vector of n (stationary) variables at time t.

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}$$

An order p in VAR implies to be regressed on its lags. Thus:

$$y_t = c + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots\ldots\ldots\ldots + \beta_p y_{t-p} + \varepsilon_t \quad (8)$$

Where:

$c = (nx1)$: Vector of constants

$\beta_j = (nxn)$: Matrix of coefficients

$\varepsilon_t = (nx1)$: Vector of white noise innovations,

$E[\varepsilon_t] = 0$ and $E[\varepsilon_t \ \varepsilon'_t] = \{^{\Omega \, if \ t=\phi}_{0 \ otherwise}$

From here, a selection of lag structure is required and some selection information criteria such as Akaike Information Criterion (AIC), Schwarz Criterion (SIC or BIC), and the Hannan-Quinn Criterion (HQ) are used to check good ε_t.

The National Bank of Rwanda adopted the VAR approach as one of the methods to model and forecast inflation in the near-term. This block produces the estimates for VAR model for each CPI component and produces estimates for the Joint VAR. The joint VAR is included in forecast averaging.

b.1.1 VAR core inflation

$$\pi_t^c = \beta_{1,j}\pi_{t-j}^c + \beta_{2,j}exr_{t-j} + \beta_{3,j}\widehat{ciea}_{t-j} + \beta_{4,j}ibr_{t-j} + c; \quad j = 1 \ to \ 4 \quad (9)$$

Core inflation VAR model includes 4 lags and the explanatory variables consist of interbank rate (ibr), nominal exchange rate (exr) while CIEA hut is the composite index of economic activities' gap (i.e. the percentage deviation of actual composite index of economic activities from its long-run/potential level) to capture domestic demand. c stands for constant term in the equation.

b.1.2 VAR food inflation

$$\pi_t^f = \beta_{1,j}\pi_{t-j}^f + \beta_{2,j}exr_{t-j} + \beta_{3,j}\widehat{gdp\_agr}_{t-j} + \beta_{4,j}\widehat{imp\_cons\_fd}_{t-j} + c; \quad j = 1 \quad (10)$$

The VAR model for food inflation is specified with a lag length of 1 for the AR term and the variables selected to explain food inflation are: agricultural GDP gap (gdp_agr hut) used as a proxy for the supply shocks: i.e. shocks due to seasonal ups and downs in food supply; nominal exchange rate (exr). US dollar is the most used among foreign currencies to import food consumed on the local market. Another variable is imported consumer food gap (imp_cons_fd hut). c is a constant term.

b.1.3 VAR energy inflation

$$\pi_t^e = \beta_{1,j}\pi_{t-j}^e + \beta_{2,j}exr_{t-j} + \beta_{3,j}imp\_energy_{t-j} + c; \quad j = 1 \ to \ 4 \quad (11)$$

In this model, energy inflstion is a function of fuel imports (imp_energy) and the nominal exchange rate (exr). is a constant term. The model has a lag length of 4.

b.1.4 Joint Vector Autoregressive (VAR)

A joint estimation of the previous three VARs is performed in order to exploit the possibility of spillover effects from one component of inflation to another. The explanatory variables include the ones included to explain VAR models for core, energy and food inflation.

b.2 Vector Autoregressive with external variable (VARX) model

The simple VARs specified and estimated so far (b.1.1 to b.1.4) are only applicable for a closed-economy since only domestic variables were taken into consideration while ignoring the influence of foreign variables that can affect inflation in one way or another. Let x^* be a foreign variable. Examples include international oil prices, output gap for the eurozone and US, used as a proxy for foreign demand.

Given that Rwanda is a small open economy, variables in may be caused/predicted by foreign variables: $x^* \rightarrow y$, e.g. international oil prices affect output and prices in Rwanda but not vice-versa. In addition to not producing oil, Rwanda's demand for oil does not exert considerable pressure on the international oil prices. This simply means that there is no bi-directional effect where Rwandan oil demand and supply conditions can affect international oil prices. From here, x^* is included in the VAR as predictors even simultaneously. There is absence of endogeneity problem since foreign variables are exogenous. This block produces the estimates for VARX model for each CPI component and produces estimates for the Joint VARX. The joint VARX is part of the forecast average. The standard nomenclature for VAR with foreign variables is VARX*. From the foregoing analysis, the VARX* is specified as follows:

$$y_t = c + \beta_1 y_{t-1} + \ldots\ldots\ldots\ldots \beta_p y_{t-p} + c_1 x_t^{\ *} + \ldots\ldots\ldots + c_{k+1} x_{t-k}^{\ *} + \varepsilon_t \tag{12}$$

$$x^* = F_1 x^*{}_{t-1} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + F_1 x^*{}_{t-q} + \xi_t \tag{13}$$

With $x^* = (x_1^{\ *}, x_2^{\ *}, \ldots\ldots\ldots x_m^{\ *})'$ -Vector of foreign exogenous variables. x^* enters VAR for simultaneously (and with lags) while does not enter a model for x^* simultaneously or in lags.

With the empirical estimations, the National Bank of Rwanda uses the below equations to forecast inflation:

b.2.1. VARX Core inflation

$$\pi_t^c = \beta_{1,j}\pi_{t-j}^c + \beta_{2,j}exr_{t-j} + \beta_{3,j}\widehat{ciea}_{t-j} + \beta_{4,j}ibr_{t-j} + c + \beta_{5,j}\widehat{USGDP}_{t-3}; \quad j = 1\ to\ 4 \tag{14}$$

This equation extends equation 9 with US Gross Domestic Product gap (USGDP hut) as an exogenous foreign demand pressures.

b2.2 VARX Food inflation

For food inflation, in equation 10 we incorporate world food prices (w_food_price) to account for external shock as some food items are imported.

$$\pi_t^f = \beta_{1,j}\pi_{t-j}^f + \beta_{2,j}exr_{t-j} + \beta_{3,j}\widehat{gdp_{agr}}_{t-j} + \beta_{4,j}\widehat{\imath mp_{cons}}_{fd_{t-j}} + c + \beta_{5,j}w_{food_{price}}{}_{t-3}; \quad j = 1 \quad (15)$$

b.2.3 Energy inflation

Similar to core and food inflation equations, the modeling of energy inflation is extended with oil prices on international market (w_oil_price) as fuel is imported.

$$\pi_t^e = \beta_{1,j}\pi_{t-j}^e + \beta_{2,j}exr_{t-j} + \beta_{3,j}imp\_energy_{t-j} + c + \beta_{4,j}w\_oil\_price_{t-3}; \quad j = 1\ to\ 4 \quad (16)$$

The joint estimation is carried out again, combining equation 14 to 16 to account for interaction of all the macroeconomic variables incorporated in these equations.

b.3 Bayesian Vector Autoregressive (BVAR) model

Bayesian methods have proven useful in the estimation of straightforward reduced-form VARs. Classical VAR methodology suffers from over parameterization. The BVAR models that are in the forecasting combination are in the spirit of Doan, (Litterman & Sims, 1986) based on the Minnesota prior, which includes the prior mean of the VAR parameters to zero, and with a prior variance depending on two hyperparameters.

BVAR includes n variables and p lags

$$y_t = c + A(L)y_{t-1} + e_t \quad ; \quad E(e_t e_t') = \varepsilon_t \quad (17)$$

The number of parameters in c and A is n(1+np). e.g., n=4, p=4: then 68 parameters ought to be estimated; with n=8, p=4: 264 parameters must be estimated. Once the sample size is over-fitted, forecasting performance reduces. Modeling and forecasting with Bayesian VAR was put in place due to the issue of dimensionality observed in the classical VAR. In doing so, some adjustments on the degree about the overall tightness hyper-parameter, the relative cross-variable weight hyper-parameter, the lag decay hyper-parameter and the exogenous variables hyper-parameter are vital to be identified.

Sims (1980) stated that even with a small system, the Bayesian approach is recommended for forecasting especially over relatively long horizons. Generally, Bayesian coefficient estimates combine information in the prior with evidence from the data. It captures changes in beliefs about parameters, priors, which means initial beliefs (e.g. before data are seen). In Bayesian estimation, posterior means new beliefs (initial beliefs + evidence from data).
We estimate a BVAR model for each CPI component and produces estimates for the Joint BVAR and BVAR with relative prices, by applying the Minnessota prior on equations 9 to 16. The values for the prior are: l1=0.1, l2=0.99, l3=1, and mu1=1; to represent the overall tightness hyper-parameter, the relative cross-variable weight hyper-parameter, the lag decay hyper-parameter, the exogenous variables hyper-parameter, respectively. In joint estimation, mu1 is zero.

Building on the research conducted on the existence of second-round effects of the non-core inflation to headline, Manishimwe (2018) demonstrated that the effect exists though it is transitory. In modelling and forecasting, a shock to one of the components, such as food inflation, may affect the changes in prices of core items. Therefore, it is paramount to model all components together and capture the effects of relative prices. In this respect a joint estimation of BVAR with relative prices models is performed and the relative prices are measures as the ratio of cpi_core over the overall cpi, cpi_food over overall cpi and cpi_energy over overall cpi for the relative price of core, food and energy respectively.

3.1.2 Forecasts performance evaluation

After specifying the models, the forecasts are done with one-step-ahead procedures using the same aforementioned equations. Once the forecasts are produced, either using a benchmark or a multivariate model, a kind of forecast performance evaluation is applied to better assess the accuracy of the forecast. Inflation forecasts need to be evaluated for a better monetary policy formulation (Clark & McCracken, 2010). With accurate forecasts, policymakers are optimistic and by having inaccurate forecasts whether they underestimate or overestimate, they incur additional costs as they mislead the decision-making process. Different methods are used to judge the forecasts performance but in this study, our focus is the forecasts accuracy and its measurements that are the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), Theil Inequality Coefficients (Theil U1 and U2).

(a) Mean Absolute Error (MAE)

The Mean Absolute Error measures the forecasts accuracy and it captures the average magnitude of the errors from multiple forecasts by ignoring their direction. The below formula weighs equally individual differences in the forecasts average.

Empirically, MAE is written as follows:

$$MAE = \frac{1}{N}\sum_{t=1}^{N} | e_{t+h/t} | \qquad (18)$$

The applied rule on the MAE is that the lower value indicates the forecast accuracy.

(b) Root Mean Square Error (RMSE)

The Root Mean Square Error is used also to diagnose the accuracy of the forecasts. It measures average magnitude of the errors like MAE. Once the Root Mean Square Errors equals to the Mean Absolute Errors, it simply means that all the errors have the same magnitude. Both RMSE and MAE fall between . The empirical specification of the Root Mean Square started from the Mean Square Error (MSE) and it is as follows:

$$MSE = \frac{1}{N}\sum_{t=1}^{N} (e_{t+h/t})^2 \qquad (19)$$

Generally, the root of MSE is taken to preserve the unit. This gives the common measure of root mean squared error. Thus:

$$RMSE = \sqrt{MSE} \tag{20}$$

From the above, the lower value, the better the forecast.

(c) Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is so called mean absolute percentage deviation and it also another measure of forecast accuracy. The formula is expressed here below:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \tag{21}$$

From the above equation, A_t refers to the actual value while F_t indicates the forecast value. As the value is expressed in terms of percentage, the above equation has to be multiplied by 100. The absolute value in this computation needs to be summed up for every forecasted point in time and divided by the number of observations N. By multiplying with 100 percent, it helps to get a percentage error. Similar to the MAE and RMSE, the lower the value signals the forecast accuracy.

(d) Symmetric Mean Absolute Percentage Error (SMAPE)

This method helps to measure the forecast accuracy based on percentage or relative error. The empirical specification is defined here below:

$$SMAPE = \frac{100\%}{N}\sum_{t=1}^{N}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \tag{22}$$

In line with the above equation, A_t represents the actual value while F_t refers to the forecast value. Under this formula, the absolute value resulted from the difference between A_t and F_t that is divided by a half of the summation of the absolute values of the actual value symbolized by A_t and the forecast value as indicated by F_t. Each value from the computation has to be summed up for every fitted point and then divided by the number of fitted points N.

(e)Theil's formula

(Theil, 1966) brought a formula to measure the forecast accuracy. This is done using the "Theil Inequality Coefficient". According to (Biemel, 1973), Theil's coefficient of inequality is considered as the statistical forecasting tool that is used to evaluate the performance of the forecasts. This measure is divided into two categories that are Theil U1 and Theil U2. Thus, the empirical specification for each is highlighted here below:

$$Theil\ \ U1 = \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2\right]^{1/2}}{\left[\frac{1}{n}\sum_{i=1}^{n}A_i^2\right]^{1/2} + \left[\frac{1}{n}\sum_{i=1}^{n}P_i^2\right]^{1/2}} \tag{23}$$

From the specification above, A_i refers to the actual observation while P_i represents the forecast value. This coefficient is denoted as *U1* in this study. Under economic forecasting, Theil appears to measure the quality of the forecast.

$$Theil\ U2 = \frac{\left[\sum_{i=1}^{n}(P_i - A_i)^2\right]^{1/2}}{\left[\sum_{i=1}^{n} A_i^2\right]^{1/2}} \tag{24}$$

Here, (A_i, P_i) indicates the observed and the predicted values.

Looking at the Theil *U1* and Theil *U2*, the difference arises from the absence or presence of the *P* term in the denominator. Having a *P* term in Theil *U1* leads the *U1* to vary between 0 and 1. For the properties, once *U=0* , there is an equality where $P_i = A_i$ for all *i* . If this materializes, there is a perfect forecast.

3.1.3. Combination of forecasts

In this study, the forecast combinations follow the work done by Kapetanios et al. (2007) on the forecast combination and the Bank of England's suite of statistical forecasting models. To them, combining different forecasts reduces forecast error. This does not necessarily mean that in all periods, forecast combination beats the benchmark models as in a very short period like one month the benchmark models may beat the combined forecasts.

In addition, the work follows also Stock and Watson (2002)'s approach that assigns equal weights to all models. The approach assumes that variances of the forecast errors are the same, the pair-wise correlations of forecast errors are the same considering that models are more than two (M>2) and the symmetric loss function, meaning that the sign and size of the forecast are disregarded.

Assume that Y_i is the variable of interest with *N*, the number of forecasting models that do not possess the perfect collinearity.

$$F_t^* = (f_{1t} + f_{2t} + \ldots\ldots\ldots\ldots f_{nt}) \tag{25}$$

From here, the simple average assigns equal weights to all forecasting models and then we have:

$$W = \frac{1}{N} \tag{26}$$

Therefore, the combined forecast becomes:

$$Y_t = F_t^* * W \tag{27}$$

In line with the above, the National Bank of Rwanda empirically uses the simple average estimation in combining forecasts from core, energy and food CPI components.

(a).Forecast combination for core inflation

$$\pi_t^{core_f} = Average\left(\pi_t^{core_{fp}}, \pi_t^{core_{fj}}\right)$$

(b).Forecast combination for energy inflation

$$\pi_t^{energy_f} = Average\left(\pi_t^{energy_{fp}}, \pi_t^{energy_{fj}}\right)$$

(c).Forecast combination for food inflation

$$\pi_t^{food_f} = Average\left(\pi_t^{food_{fp}}, \pi_t^{food_{fj}}\right)$$

As early mentioned in the methodology, only the forecasts from Joint models and Joint models with relative prices (*fp**) such as Joint BVAR and Joint BVARX are the ones contributing to the simple averaging scheme to make forecast combination for each CPI component. Such models were selected as they outperformed better than benchmark models in terms of having minimum root mean square errors. After having the forecasts of each component, we generate the forecast of headline inflation.

The process is done by computing total consumer price index as a weighted sum of food consumer price index (with 0.32233 weight in total consumer price index), energy consumer price index (with 0.0675980 weight in total consumer price index) and core consumer price index (with 0.610072 weight in total consumer price index). The values of these indices are obtained by considering the respective inflation in the subsequent period. Regarding the headline inflation, it is obtained by calculating the percentage change of total consumer price index.

3.2. Variables selection and data

3.2.1. Selection of variables and definitions

The choice of variables was guided by the theories on inflation, macroeconomic fundamentals of the selected context that is the Rwandan economy, previous works on inflation dynamics and forecasting models as well as expert judgments.

Among the theories on determinants of inflation, the study embraced the Keynesian theory that explains demand and cost pull inflation (Doguwa & Alade, 2013) as well as the purchasing power parity theory. In addition, some empirical works on modeling and forecasting inflation focusing on Rwandan context were used as a reference to select variables and include the study done by (Gichondo & Kimenyi, 2012; Ruzima & Veerachamy, 2015; Manishimwe, 2018; Mwenese & Kwizera, 2018 and Maniraguha, et al., 2019). Here below is the list of both endogenous and exogenous variables that are used in this study:

Table 1: Endogenous and exogenous variables

Variables	Definition	Motivation
Endogenous		
cpi_core	Core consumer price index that excludes food items	Main component of total consumer price index
cpi_food	Food consumer price index	Main component of total consumer price index
cpi_energy	Energy consumer price index	Main component of total consumer price index
Exogenous		
IR_IB	Interbank rate	To measure the effects of monetary policy on core inflation
rwf_usd	Nominal Exchange Rate	To measure the influence of exchange rate on core, food and energy inflation
CIEA	Composite Index of Economic Activities	To capture pressures on core inflation from domestic demand
Imp_cons_food	Imported consumer food	To measure the magnitude of imported food items on food inflation
cl_gdp_agr	Agricultural output gap	To capture the effects of the proxy for the supply shocks: i.e. shocks due to seasonal ups and downs in food supply agricultural products on food inflation and even on core food inflation.
Imp_energy	Imported energy	To know the magnitude of fuels costs on energy inflation
Y_US_GAP	US output gap.	It helps to capture pressures from external side on core inflation
Wfood	World food prices	To capture the effects of foreign food prices
Oil	International oil prices	To detect the effects of international oil prices on energy inflation
C	**Constant term**	

3.2.2 Data

This study uses the time series data spanning from 2012M02 to 2020M09. The sample size was chosen to have a balanced dataset, as the Composite Index of Economic Activities (CIEA) was available since 2012M02. As it is done under the Quarterly Projections Model (QPM), data were collected from different sectors such as balance of payment, global, inflation, real and monetary sectors and most of the data were seasonally adjusted through matlab codes except interest rate and they were imported in Eviews program for further transformation, estimation and forecasting purposes. Some data are published by the National Institute of Statistics of Rwanda like Gross Domestic Product and Consumer Price Indices, while the other remaining variables are computed and published by the National Bank of Rwanda via the Statistics and Monetary Policy Departments.

Yet again, data that are in raw frequency terms (quarterly data) were converted into high-frequency terms (monthly data) using Dentons (1971)'s algorithm. This benchmarking approach of converting quarterly data in monthly frequency brings consistency in data though it may possess some limitations like the generation of measurement errors and the lack of theoretical foundation to support some correlations. Like other model specifications and regression analysis, data that come in the analysis are differenced except interest rate to satisfy the stationarity aspect.

4. EMPIRICAL RESULTS

This section describes empirical findings from forecast evaluation, the one-step-ahead forecasts as well as forecast combinations applied on disaggregated headline CPI components.

The study estimates the main components of total headline CPI, that is core that excludes all food items from headline CPI, food and energy. The estimation covers the 2012M02-2020M06 sample while the out-of-sample forecasts are done using one step ahead forecast technique and are for the period 2020M10-2020M12. The three (3) components of headline inflation are estimated and forecasted using benchmark models (ARMA and Random Walk), VARs, VARX, Joint VAR, BVAR, BVARX, Joint BVAR, and Joint BVARX with relative prices. Before combining the aforementioned forecasting models to get one single forecast of headline inflation for three (3) months ahead, in-sample forecasting evaluation was done using econometric techniques that detect the forecast accuracy per each model.

4.1. Forecasts evaluation

4.1.1-Forecast evaluation and comparison for core inflation

Except Random Walk and ARMA specification models, core inflation was estimated by considering its past behavior, domestic factors that are theoretically known to affect core inflation (exchange rate, composite index of economic activity, interbank), coupled with external factors that include US GDP gap. The forecasts for core inflation were evaluated mainly by looking at the level of accuracy.

From the results presented in table 2 below, it is observed that the best in-sample forecasts of core inflation are given by the Joint Bayesian VAR with relative prices, Joint Bayesian VAR and VARX as well as Joint VAR and VARX. However, the benchmark models that are Random Walk and ARMA models remain the worst models for the chosen sample size. Generally, the majority of tests indicated that core inflation forecasts have the minimum errors once the core inflation model is estimated with multivariate variables (domestic and external side), adding to consideration of relative prices. The literature on these techniques include (Chai & Draxler, 2014; Willmott, et al., 2009). Basing on these findings, it does not necessarily mean that Joint Bayesian and Joint VAR models are the best for each horizon as this depends on the sample size taken into consideration. Though the benchmark models are beaten in more than one month, they are likely to provide good forecasts in one month ahead.

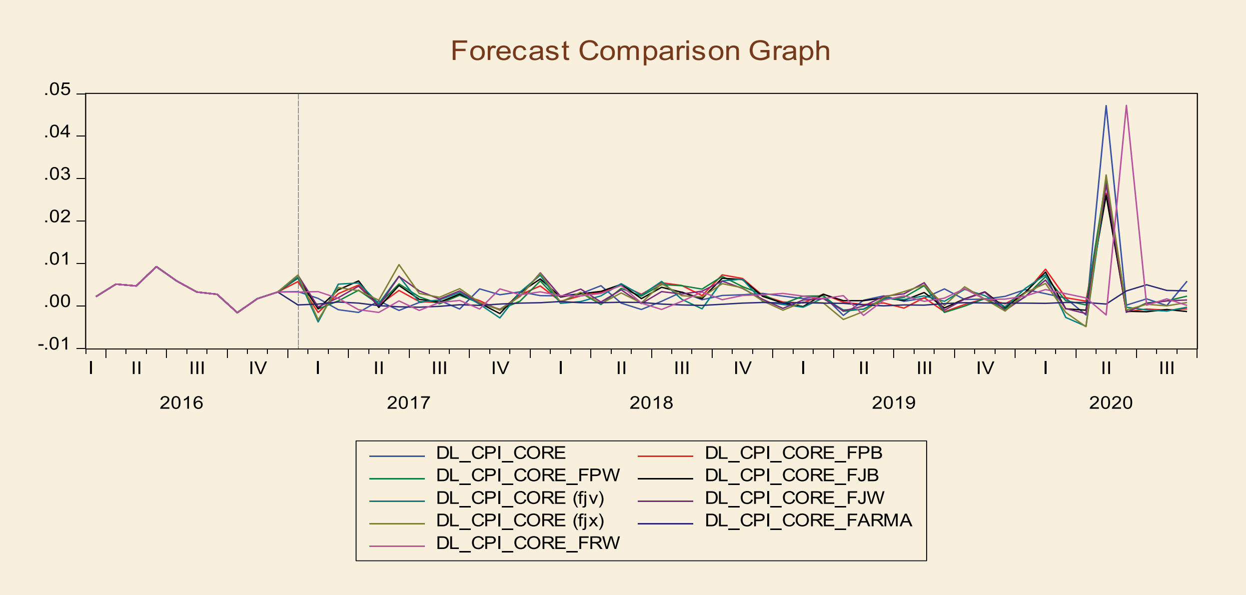
Table 2: Forecasting models and forecast evaluation for DL_CORE_CPI inflation

Forecasting models for CORE CPI inflation	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
DL_CPI_CORE_FPB[32]	0.004236	0.002744	186.9442	112.1905	0.339084	1.102734
DL_CPI_CORE_FPW[33]	0.003764	0.002642	183.1107	113.0124	0.293386	0.620167
DL_CPI_CORE_FJB[34]	0.004352	0.002757	182.9266	108.6902	0.347782	1.206719
DL_CPI_CORE_FJV[35]	0.004321	0.003092	205.3127	118.8214	0.332942	1.047722
DL_CPI_CORE_FJW[36]	0.003996	0.002703	188.6899	110.6122	0.31278	0.756755
DL_CPI_CORE_FJX[37]	0.004069	0.002923	176.6577	110.996	0.30839	0.851925
DL_CPI_CORE_FARMA[38]	0.007304	0.002971	228.3668	141.7388	0.835129	0.590527
DL_CPI_CORE_FRW[39]	0.010381	0.003752	842.0307	99.01158	0.69919	0.998955

Source: Authors

[32] FPB symbolizes the Joint Bayesian VAR with relative prices model
[33] FPW refers to the Joint Bayesian VARX with relative prices model
[34] FJB means Joint Bayesian VAR model
[35] FJV refers to the Joint VAR model
[36] FJW, reflects Joint Bayesian VARX model
[37] FJX, symbolizes Joint VARX model
[38] FARMA, represents an Autoregressive Moving Average model
[39] FRW, represents the Random Walk

Figure 1: In sample forecasts comparison for core inflation



4.1.2 Forecast evaluation and comparison for food inflation

In line with food inflation forecasts evaluation, firstly the drivers of food inflation were estimated. The drivers included the factors from domestic and external sides. Those factors include those from domestic side like the past trend of food inflation or lags, exchange rate, agriculture GDP gap, import consumer foods, while the external factors include world food price index. Like core inflation forecasts, food inflation forecasts within the sample are distributed among different models as indicated in table-2.

With the results concerning the food forecasts evaluation as shown in table 3 below, the Joint Bayesian VAR with relative prices, Joint Bayesian VAR and VARX as well as Joint VAR and VARX outperform the benchmark models for the specified sample size. This implies that the majority of tests indicated that food inflation forecasts have the minimum errors when the estimated model reflects multivariate variables either allowing for only domestic and external variables, as well as taking into account relative prices.

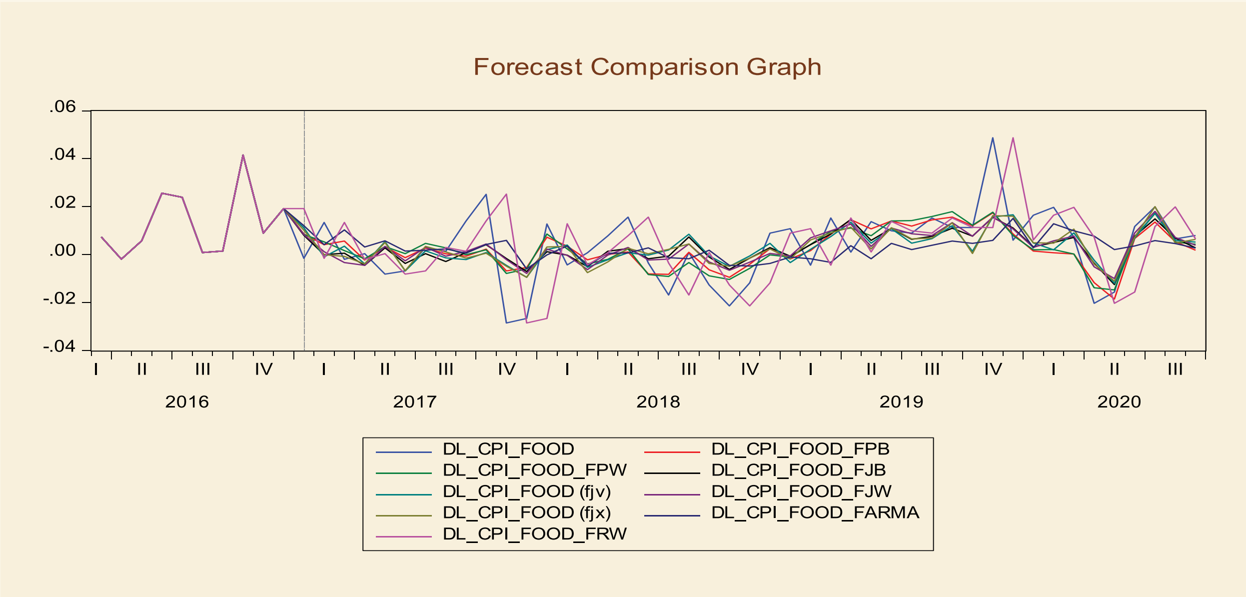
The literature on statistical and multivariate models as well as forecast combinations includes the work done by (Kapetanios, et al., 2007). On one hand, once the forecasts from different models are combined, forecast errors are reduced, meaning that joint models may perform better than statistical/ univariate models for some periods. On the other hand, such models may perform poorly in some periods, which means, it does not necessarily mean that in all periods, forecast combination beats the benchmark models. For example, in very short periods like one month, forecasts from the statistical models may beat the combined forecasts.

Table 3: Forecasting models and forecast evaluation for DL_FOOD_CPI inflation

Forecasting models for FOOD CPI inflation	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
DL_CPI_FOOD_FPB	0.010662	0.008281	131.027	108.1454	0.461313	0.95975
DL_CPI_FOOD_FPW	0.010284	0.008004	132.7738	109.2704	0.43778	0.97635
DL_CPI_FOOD_FJB	0.011394	0.00899	165.5474	120.3245	0.527962	1.018074
DL_CPI_FOOD_FJV	0.012122	0.009803	176.7952	127.6308	0.549535	1.225311
DL_CPI_FOOD_FJW	0.011106	0.008627	151.7474	110.1096	0.506282	1.047285
DL_CPI_FOOD_FJX	0.011734	0.009351	158.8974	116.8897	0.526935	1.194126
DL_CPI_FOOD_FARMA	0.014166	0.011403	153.9679	142.5238	0.697772	1.266915
DL_CPI_FOOD_FRW	0.017769	0.013362	265.0033	123.5013	0.593087	1.000217

Source: Authors

Figure 2: In sample forecasts comparison for core inflation



4.1.3 Forecast evaluation and comparison for energy inflation

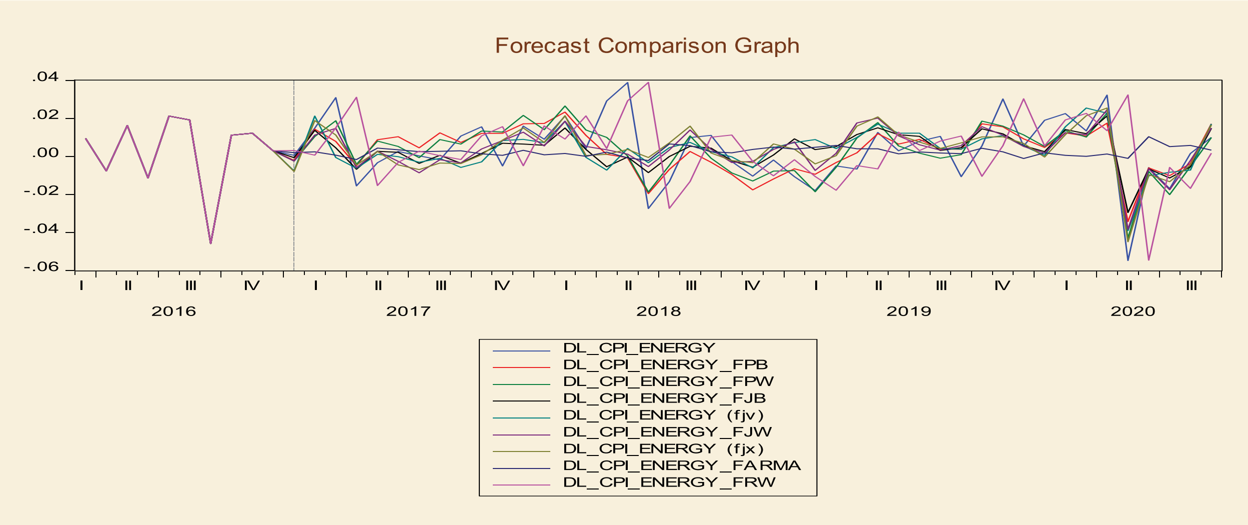
Just like core and food inflation forecasts, energy inflation forecasts were produced by estimating energy equations using different techniques as applied to other components of headline CPI inflation. The dissimilarity occurs in the drivers where the energy equation centered on the past behavior of energy, exchange rate, imported energy (fuels), coupled with external factor that is international oil prices. Table-4 below presents the results of energy forecasts evaluation. The Root Mean Square Errors (RMSE) which is one of the indicators of forecast accuracy shows that the Joint models (BVAR and VARs) perform better than statistical models for the selected sample size.

This implies that the mainstream tests indicated that energy inflation forecasts have the minimum errors when the model estimation reflects multivariate specifications by either allowing for only domestic variables, as well as taking into account relative prices.

Table 4: Forecasting models and forecast evaluation for DL_ENERGY_CPI

Forecasting models for ENERGY CPI inflation	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
DL_CPI_ENERGY_FPB	0.011824	0.009026	157.7648	93.99985	0.400607	1.627612
DL_CPI_ENERGY_FPW	0.010443	0.007992	123.0005	89.61873	0.335793	1.247737
DL_CPI_ENERGY_FJB	0.013472	0.009972	124.9751	110.3745	0.499194	0.655591
DL_CPI_ENERGY_FJV	0.014076	0.010933	161.61	122.6768	0.486532	0.724202
DL_CPI_ENERGY_FJW	0.012322	0.009398	167.5073	104.1018	0.427676	0.606697
DL_CPI_ENERGY_FJX	0.012864	0.010103	189.2587	114.5717	0.431132	0.625985
DL_CPI_ENERGY_FARMA	0.017442	0.01362	136.3009	160.6203	0.833154	1.026755
DL_CPI_ENERGY_FRW	0.023274	0.016408	220.99	129.2813	0.667385	0.995966

Figure 3: In sample forecasts comparison for food inflation



Since the accurate in-sample forecasts are provided by core, food and energy inflation by using the Bayesian VAR and VAR models, coupled with relative prices, these results are in line with the study done in Nigeria by (Kelikume & Salami, 2014).

By comparing VAR and ARIMA models in terms of forecasting performance on monthly CPIs data spanning from 2003M01 to 2012M06, their results showed that VAR model specification outperformed the ARIMA model, looking at the minimum Root Square Errors (RMSEs). A similar study was conducted in South Africa (Gupta & Kabundi, 2008) but dynamic factor models were found to predict macroeconomic indicators including headline inflation better than New Keynesian DSGE models, basing on the minimum RMSE. Still in Nigeria, Doguwa and Alade, (2013) compared forecasts of headline inflation by separating its main components, which is food and core inflation. Following the RMSE test, the SARIMA was found appropriate to produce good forecasts of food inflation for less than 10 months ahead while SARIMAX was found powerful for eleven and twelve months ahead. Furthermore, the RMSE as a test of forecast accuracy was applied in Slovenia. The study conducted by Krušec (2007) confirmed that factor models performed better than AR benchmark model and somehow similar to the VAR models on all inflation components including energy inflation, industrial goods inflation, services inflation, processed food and non-processed food inflation.

4.2 The one-step-ahead forecasts, forecast combinations

In this study, out of sample monthly forecasts were done for the horizon starting from 2020M10 to 2020M12. Such forecasts were produced using a one-step-ahead forecast technique. According to Stock and Watson (2002), this forecasting technique directly provides rapid results. This was applied to each component of headline inflation as early mentioned. However, the forecast combination took into account only joint models following their performance and theory. The joint models constructed on multivariate basis were chosen to be part of forecast combination due to their forecasts' accuracy in comparison with benchmark models (Random walk and ARMA models). Inflation was calculated in different frequencies, namely, Y-o-Y, Q-o-Q and M-o-M and for core, food, energy as well as the total headline inflation, respectively. Headline inflation in similar frequencies was calculated via aggregation of the components using their respective weights.

To get the forecasts of the three (3) components, a simple average of joint models technique was applied. Once the inflation forecasts are determined at each month ahead for every component, headline inflation is calculated using the weighted average formula. More details on forecasting combination are in Section III. The results from the out-of-sample forecasts using one step ahead revealed that headline inflation (Y-o-Y) would record a slight decline and was expected to evolve around 8.32percent in October 2020. It will decline further in November at 6.73 percent while picking to 7.02 percent in December 2020.

Additionally, the forecasts from the combined models (core, food and energy) point out that headline inflation will evolve around 7.36 percent in 2020Q4.In line with the quarterly projections model (QPM) operation during MPC processes at the National Bank of Rwanda, short-term inflation forecasts are put in QPM as the starting point of 8 quarters ahead. In doing so, expert judgments must be incorporated after producing these aforementioned short-term models. These judgments are based on results of the food price expectation survey that is normally done on quarterly basis. Besides, judgments on the forecasts resulted from experience, theories, or previous research findings, coupled with macroeconomic fundamentals of Rwanda. Here below are the results of headline inflation forecasts from combined models as highlighted in section III and these do not include expert judgments. They are only forecasts from the model.

Table 5: Short-term inflation forecasts (2020M10-2020M12)

MoM	Energy	Food	Core	Headline
2020m10	-0.3267	0.678963	0.112703	0.291430
2020m11	2.494274	1.467836	-0.43028	0.465576
2020m12	2.076798	0.812528	0.415580	0.675511
YoY	Energy	Food	Core	Headline
2020m10	5.378110	11.53505	6.641208	8.327312
2020m11	4.847986	8.125016	6.049295	6.733185
2020m12	6.381244	8.329920	6.237232	7.020108

Quarterly forecasts (2020Q4)	Energy	Food	Core	Headline
QoQ	2.752378	2.677551	0.356346	1.371457
YoY	5.535780	9.329996	6.309245	7.360202

Source: Authors

Figure 5: Out of sample for core

Figure 6: Out of sample for food

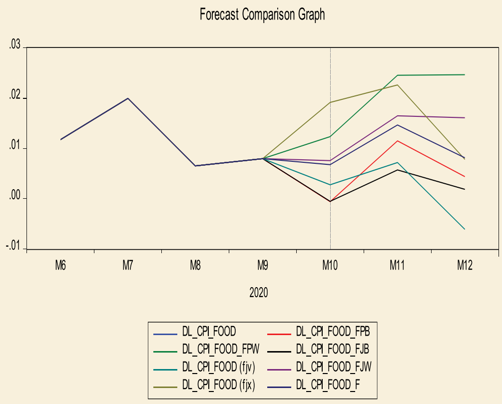
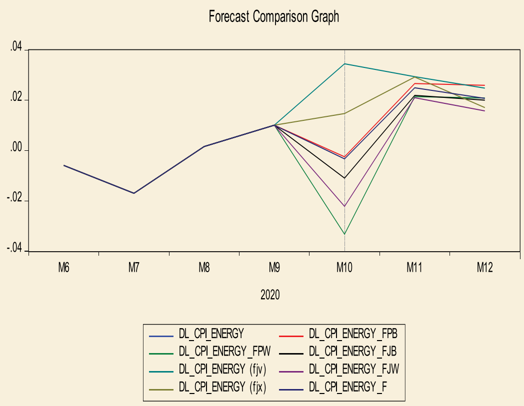


Figure 5: Out of sample for energy



5. CONCLUSION AND POLICY RECOMMENDATIONS

Near-term forecasting for inflation plays a crucial role in monetary policy formulation. At the National Bank of Rwanda, the NTF process helps in the assessment of current developments and initial conditions, provides the starting point to the QPM model as well as the prediction of variables not included in the core model. This study produces short-term inflation forecast for the National Bank of Rwanda by using different econometrics tools. Building on the recent literature that supports forecasting models and forecast combination, three CPIs components (core that excludes food items, food and energy) were estimated, forecasted and then averaged to get one single headline inflation forecast. The rationale of producing short-term inflation forecast in a disaggregated manner and combining the outcome resulted from the advantage of forecast combination in policymaking. Inflation forecast resulting from combined models helps to avoid the confusion associated to what forecast to pick as accurate. Moreover, results from combined forecast are robust as the approach use a relatively larger information set.

The estimation of the three (3) components of headline inflation was performed using benchmark models (ARMA and Random Walk) and multivariate models (VARs, VARX, Joint VAR, BVAR, BVARX, Joint BVAR, and Joint BVARX with relative prices). Their forecasts were combined to produce a single forecast of headline inflation. This was done basing on the minimum forecast errors. The results indicated that the forecasting combination done using the simple averaging scheme technique reduces forecast error compared to the individual model forecasts and these findings are in line with works done by Kapetanios et al. (2007) for Bank of England, and Bjørnland et al. (2008) for Norges Bank.

However, the benchmark or statistical models remain the worst models over the sample selected for forecast evaluation as indicated by the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE) and Theil Inequality Coefficients (Theil U1 and U2). Once the forecasts of inflation on core, food and energy are found, a simple average of joint models technique was applied. The forecasts are determined at each month ahead for every component and headline inflation was calculated using weighted average formula. Thus, the results revealed that headline inflation would record a pick (8.32percent) in October 2020 and slightly decelerating in November 2020 to 6.73 percent. Unlike November 2020, the findings showed that headline inflation is expected to pick at 7.02 percent in December 2020.These forecasts project quarterly headline (Y-o-Y) to be around 7.36 percent in 2020Q4.

The contribution of this study resides not only in the disaggregation of the total headline components but also in incorporating models of joint VAR, Joint BVAR and Joint relative prices that at the end become part of the forecasting combination. The capability of the joint models relies on its flexibility which is to allow more information helping to capture inflation dynamics as this includes both domestic and external drivers of headline inflation.

From the perspectives of short-term inflation forecasting models and forecast combination at the National Bank of Rwanda, the study recommends the continuous use of combined multivariate models for short-term headline inflation forecasts as well as to consider these forecasts to feed the quarterly projections model (QPM) as they include more information on economic dynamics than statistical models. In addition, like Box and Jenkins (1976) stated "all models are wrong but some are useful", it is paramount to improve the aforementioned forecasts by incorporating expert judgments before their utilization in monthly inflation reports or medium-term forecasting models. In line with the current forward-looking framework adopted by the National Bank of Rwanda, findings from this paper have an important implication in terms of monetary policy orientation. They inform monetary policy authorities of the future trend of inflation within a horizon of one to 3 months ahead. By having an insight on such trajectory, policymakers may react accordingly towards price stability, which is an objective of the central bank.

REFERENCES

Atiya, F., 2020. Why does forecast combination work so well?. International Journal of Forecasting, 36(1), pp. 197-200.

Bañbura, M., Giannone, D. & Reichlin, L., 2010. Nowcasting, Frankfurt am Main: European Central Bank (ECB).

Bates, J. & Granger, C., 1969. The combination of forecasts. Operational Research Quaterly, 20(4), pp. 451-468.

Bernanke, S., 2005. Inflation in Latin America.A New Era?. Stanford, California, Stanford Institute for Economic Policy Research Economic Summit.

Biemel, F., 1973. Theil's Forecasts Accuracy Coefficient:A clarification. Journal of Marketing Research, Volume 10, pp. 444-446.

Bjørnland, C., Jore, A., Smith, C. & Thorsrud, L., 2008. Improving and evaluating short term forecasts at the Norges Bank, s.l.: Norges Bank.

Chai, T. & Draxler, R., 2014. Root Mean Square Error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. Geoscientific Model Development, Volume 7, pp. 1247-1250.

Clark, E. & McCracken, W., 2010. Averaging forecasts from VARs with uncertain instabilities. Journal of Applied Econometrics, 25(1), pp. 5-29.

Diebold, X. & Lopez, A., 1996. Forecast evaluation and combinations, Massachusetts: National Bureau of Economic Research.

Diebold, X. & Mariano, S., 2002. Comparing predictive accuracy. Journal of Business and Economic Statistics, 20(1), pp. 134-144.

Doguwa, I. & Alade, O. S., 2013. Short-term Inflation Forecasting Models for Nigeria. CBN Journal of Applied Statistics , 4(2).

Friedman, M. & Bordo, D., 2006. The Quantity Theory of Money: A Restatement. In: Routledge, ed. The Optimum Quantity of Money.Taylor & Francis, p. 18.

Gichondo, A. & Kimenyi, V., 2012. Modelling inflation in Rwanda. BNR Economic Review, Volume 05, pp. 57-75.

Granger, W. & Ramanathan, R., 1984. Improved methods of forecasting. Journal of forecasting, 3(2), pp. 197-204.

Grasa, A., 1990. Econometric Model Selection:A New Approach. 1989, Springer.

Gupta, R. & Kabundi, A., 2008. A Dynamic Factor Model for Forecasting Macroeconomic Variables in South Africa, Pretoria: Department of Economics Working Paper Series.

Gupta, R. et al., 2012. Sequancing and analysis of a South Asian-Indian personal genome. BMC Genomics.

Harvey, C., 1989. Forecasting,Structural Time Series Models and the Kalman Filter. 1 ed. New York: Cambridge University Press.

Hsiao, C. & Wan, K., 2014. Is there an optimal forecast combination?. Journal of Econometrics , 178(2), pp. 294-309.

Humphrey, M., 1977. On Cost Push Theories in the pre-war Monetary Literature,Federal Reserve Bank of Richmond, Federal Reserve Bank of Richmond.

Jenkins, M., Box, E. P. G., Reinse, C. I. & Ljung, M., 2016. Time Series Analysis:Forecasting and Control. 5 ed. New Jersey.: John Wiley & Sons, Inc., Hoboken.

Kapetanios, G., Labhard, V. & Price, S., 2007. Forecast combination and the Bank of England's suite of statistical forecasting models, QueenMary: Bank of England.

Kavila, W. & Roux, L., 2017. The reaction of inflation to macroeconomic shocks:The case of Zimbabwe (2009-2012), Economic Research Southern Africa.

Kelikume, I. & Salami, A., 2014. Time Series Modelling and Forecasting Inflation:Evidence from Nigeria. The International Journal of Business and Finance Research, 8(2), pp. 91-98.

Khan, S. & Schimmelpfennig, A., 2006. Inflation in Pakistan:Money or Wheat?, Washington DC: International Monetary Fund.

Krušec, D., 2007. Short term inflation projections for Slovenija:comparing factor models with ar and var models, Frankurt am Main: European Central Bank.

Lack, C., 2006. Forecasting Suiss Inflation using VAR models, Zurich: Swiss National Bank Economic Studies.

Lars, E., 2000. How should Monetary Policy be conducted in an era of price stability?Working Paper 7516., Massachusetts Avenue Cambridge,MA 02138: National Bureau of Economic Research 1050.

Litterman, B. & Sims, A., 1986. Forecasting with Bayesian Vector autoregressions-five years of experience. Journal of Business and Statistics, Volume 4, pp. 25-38.

Lucas, R., 1976. Econometric Policy Evaluation: Acritique. Carnegie-Rochester Conference Series on Public Policy, 1(1), pp. 19-46.

Maniraguha, F., Karangwa, M. & Bagabe, J., 2019. Estimating time-varying volatility in consumer prices in Rwanda:Application of GARCH models. BNR Economic Review, Volume 14, pp. 53-71.

Manishimwe, C., 2018. Second round effect of food and energy prices to inflation in Rwanda.. BNR Economic Review, Volume 12 , pp. 127-139.

Montero-Manso, P., Athanasopoulos, G., Hyndman, J. & Talagala, S., 2018. FFORMA:Feature-based forecast model averaging, Monash City: Monash Business School.

Muth, F., 1961. Rational Expectations and Theory of Price Movements. Econometrica, 29(3), pp. 315-335.

Mwenese, B. & Kwizera, A., 2018. Modelling and forecasting inflation dynamics in Rwanda. BNR Economic Review, Volume 13, pp. 45-64.

Nyoni, T. & Solomon, N., 2018. Inflation modelling in Nigeria:An Application of ARMA,ARIMA and GARCH Models. MPRA.

Öðünç, F. et al., 2013. Short term inflation forecasting models for Turkey and a forecast combination analysis. Economic Modelling, Volume 33, pp. 312-325.

Robert, E. & Sargent, J., 1979. After Keynesian Macroeconomics. Quarterly Economic Review, 3(2).

Rummel, O., 2015. Economic modelling and forecasting, s.l.: s.n.

Ruzima, M. & Veerachamy, P., 2015. A study on Determinants of Inflation in Rwanda from 1970-2013. International Journal of Management and Development Studies, 4(4), pp. 390-401.

Sims, A., 1980. Macroeconomics and reality. Econometrica, 48(1), pp. 1-48.

Smithies, A., 1942. The Behavior of Money,National Income under Inflationary Conditions. Quarterly Journal of Economics,57(1), pp. 113-128.

Smith, J. & Wallis, F., 2009. A simple explanation of the forecast combination puzzle. Oxford Bulletin of Economics and Statistics, 71(3), pp. 331-355.

Spencer, E. D., 1993. Developing a Bayesian Vector Autoregression Forecasting model.. International Journal of Forecasting,Elsevier, 9(3), pp. 407-421.

Stock, H. & Watson, W., 2002. Macroeconomic Forecasting using diffusion indexes. Journal of Business and Economic Statistics, 20(2), pp. 147-162.

Theil, H., 1966. Applied Econometric Forecasts.. Amsterdam-North Holland.1966. Theory,Birmingham.

Timmermann, A., 2006. Forecast combinations. Handbook of Economic Forecasting, Volume 1, pp. 135-196.

Willmott, J., Matsuura, K. & Robeson, M., 2009. Ambiguities inherent in sums-of-squares-based error statistics. Atmospheric Environment, 43(3), pp. 749-752.

NATIONAL BANK OF RWANDA
BANKI NKURU Y'U RWANDA

NATIONAL BANK OF RWANDA, KN 6 AV.4
P.O.Box: 531 Kigali Rwanda
Tel: (+250) 788 199 000 Email: info@bnr.rw
Swiftcode: BNRWRWRW | Twitter: @CentralBankRw
Website: www.bnr.rw