ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Efficiency of Commercial Banks in Ethiopia: A Stochastic Frontier Analysis

By Sefiager Alem

A Thesis Submitted to the Graduate Studies of Addis Ababa University in partial fulfillment of the Requirements for the Degree of Master of Science in Financial Economics

Ausgust 2024

Addis Ababa Ethiopia

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Efficiency of Commercial Banks in Ethiopia: A Stochastic Frontier Analysis

By Sefiager Alem

Sisay Regassa ______________________ Sisay R.

Advisor | Signature

Zerayehu Sime ______________________

Examiner | Signature

Dr. Etsubdink Sileshi ______________________

Examiner | Signature

DECLARATION

I, **Sefiager Alem Abtew,** do hereby declare to Addis Ababa University School of Graduate Studies that this dissertation paper is a product of my original research work, and it has not been submitted to any other university for any academic degree. Materials and information other than my own are dually acknowledged.

Name: **Sefiager Alem**

Signature: ______________

Date of Submission: ______________

Abstract

The study investigates the cost efficiency and determinants of cost efficiency among Ethiopian banks over the period 1998 -2021, by employing a translog stochastic cost frontier method. Quality of loans and financial capital are accounted for in the cost function and their impact on cost efficiency level is evaluated. The result shows that the inefficiency over the period averaged 26.88 percent. For individual banks the cost inefficiency ranges from 16.64 percent to 38.28 percent of total costs. Larger and older banks have better efficiency than smaller and younger banks. The efficiency of banks over time tends to decrease until 2020 with a slight increase in 2021. The result also shows that persistent inefficiencies that relate to operation model, structure, location and other structural factors are more important than time varying factors. In addition, over the study period banks did not show technical progress instead, though very small, there seems to be a technical regress. Finally, non-core deposits to financial assets ratio and total assets of banks are negatively associated with the transient inefficiency of banks whereas share of non- core deposits to total deposits, age, ratio of other earning assets to deposits ratio, and intermediation ratio have a positive association with transient inefficiency. Generally, smaller banks are more flexible and suited to address time varying conditions than the bigger ones.

Acknowledgement

I would like to express my deepest gratitude to all those who have supported me throughout the course of this thesis. Words can not express my gratitude to him my advisor, a dear friend and a brother Sisay Regassa (PhD) for his feedback, patience and commitment. You did not just provided invaluable comments that greatly improved the thesis but also understood me very well in all the difficult times I have been through all along. I am glad to have taken this ride with a person like you. To this I am profoundly grateful. God bless you and your family!

Most importantly, I want to express my profound gratitude to my wife, Tigist Endeshaw. Your unwavering support, boundless patience, and understanding have been my cornerstone throughout this endeavor. Your sacrifices, encouragement, and belief in me have been a constant source of strength and motivation. This thesis is as much yours as it is mine.

Contents

List of Acronyms

AB – Awash Bank

AbB – Abay Bank

AdIB – Addis International Bank

BIB – Birhan International Bank

BOA – Bank of Abyssinia

BuB - Bunna Bank

CBE – Commercial Bank of Ethiopia

CPI – Consumers Price Index

Coop – Cooperative Bank of Oromia

DB – Dashen Bank

DEA – Data Envelopment Analysis

DFA – Distribution Free Analysis

EB – Enat Bank

FDH – Free Disposal Hull

GBE – Global Bank Ethiopia

IMF – International Monetary Fund

LB – Lion Bank

NBE – National Bank of Ethiopia

NIB – Nib International Bank

NPL – Non Performing Loans

OIB – Oromoia International Bank

SSA – Sub Saharan Africa

SFA – Stochastic Frontiers Analysis

TFA – Thick Frontier Approach

WDI – World Development Indicators

UB – United Bank

WB – Wogagen Bank

ZB – Zemen Bank

List of Figures

List of Tables

Chapter One: Introduction

1.1. Introduction

The financial system is an important element of any economy. Through a network of several structures and institutions, the financial system channels resources to their utmost use. The development and stability of a country‟s economy depends on the performance of the financial system. Banks are a crucial part of the financial system. Through the intermediation process, banks provide liquidity, payments, and safekeeping services for depositors and provision of funds for investment and working capital for nonfinancial industries (Berger and Mester, 1997). Through this process they ameliorate information asymmetry between borrowers and lenders, manage risks and channel funds, that otherwise would have been idle, to profitable investments. In a nutshell, banks allocate resources, spread risk and through the provision of liquidity ensure the smooth flow of resources within an economy. Therefore, problems faced by banks affect the whole economic system substantially.

The implications of bank inefficiency are manifold. Inefficient banks increase the overall intermediation cost of an economy that slows down economic growth. This makes bank efficiency a socially optimal target since it reduces the cost of financial intermediation (Resti, 1997). Inefficient banks could make sub optimal allocation of funds and lower profitability. Efficiency of banks is not associated only with better performance of banks but also with avoiding failure and ensuring resilience. The rate of failure by banks with low efficiency is higher than those with higher efficiency (Berger and Humphrey, 1997). Scores of researches have evidenced that failing banks during the 1980s banking crisis are placed far from the best practice frontier (Berger and Humphrey, 1992; Berger and De Young, 1997). Thus, efficiency of banks can be used to detect early problems of the banking system for regulators (Mester, 1996).

The Ethiopian financial system underdeveloped; the financial development index for 2021 is 11.55% (IMF, 2021). Due to the absence of secondary capital markes, now underway, and fledgling and inchoate financial market, with a financial markets depth index of 0.023 (IMF, 2021); banks play a pivotal role in the economic system of the country. Banks are also the primary financial institutions in terms of number, size and volume of operation. Banks account for 96.5 percent and 96.3 percent of total financial sector assets by 2022 and 2023 respectively (NBE, 2024). The formal economic system of the country is operated mainly through banks. Banks are almost the only source of financing for businesses and the government. Consequently, the performance of the financial system is highly dependent on the performance of the banking system. Therefore, inefficiency in the banking system begets severe implications on the overall economy of a country.

Until 1995 the Ethiopian banking sector was wholly operated by government owned banks. The first private bank was established in 1995 following the bank licensing directive SBB/01/94. Since then about thirty private banks are established and currently, there are thirty two commercial banks in Ethiopia. The number of commercial banks in the country is increasing fast recently; ten of the twenty seven banks have been established in 2021. Inspite of all this increase in the number of private banks, the Ethiopian banking system is dominated by state owned banks. By the end of 2022, state owned banks account 52.3 percent of the deposits and 53.3 percent of loans (NBE, 2022).

In an effort to further reform the sector, the government is allowing foreign banks operations under certain conditions. The combined outcome of increase in the number of banks and entry of foreign banks would be intensified competition and squeeze in profit margins. Efficiency becomes crucial to remain profitable. Nonetheless, there is limited research on the efficiency of commercial banks in Ethiopia. The few studies conducted in Ethiopia used limited observations and predominantly used non-parametric approach. Therefore, this study strives to measure the efficiency level of commercial banks in Ethiopia separating persistent structural inefficiency from transitory inefficiencies, and the impact of technical progress. Besides, the study attempts to identify the correlates of inefficiency in the Ethiopian commercial banks.

1.2. Statement of the Problem

Following development of efficiency measurement methodology by Farell (1957), efficiency research has been conducted in many sectors. Banking is not an exception; banking efficiency studies have been conducted since the early 1960s. Driven by increase in competition, following the deregulation of the banking sector in many countries, the resulting reduction in bank profitability, and subsequent crisis in several banks; bank efficiency research has proliferated in

1980‟s and 1990‟s. Yet, till the present time, the dynamics and differences in the development level of banks across the world have made banking efficiency studies to remain appealing research area. Predominant number of these studies, however, is conducted in developed countries of Europe and North America (Mosunda, 2016). In particular, bank efficiency research remains very limited in Sub-Saharan Africa (SSA) (Mosunda, 2016).

In order to liberalize the banking sector the National Bank of Ethiopia issued directive SBB/01/94 that allows private investors to establish banks. Subsequent to this directive the first private bank was established in 1995. Since then thirty private banks has been established. The engendering of private banks was expected to improve the quality of bank services, to create new financial services and lower costs (Harvey, 1996). Efficiency enhancement was, thus, one of the envisioned effects of the banking sector reform. However, much of the extant data on the banking sector imply a limited success to meet the expectation of the bank reform with regard to efficiency. In fact the overall increasing trend in the interest rate spread and bank lending rates with a more or less constant non-performing loans rate could probably signal a decline in the efficiency of banks. The interest spread in the Ethiopian banking sector, except a two year spike in 2015 and 2016, has a generally increasing trend. By 2021, the interest rate spread in the Ethiopian banking sector has increased to 7.3 percent (NBE, 2021) from its level of 6.8 percent in 2011. The bank lending rate has also been increasing over time; it increased from 12.25 percent in 2011 to 11.88 percent in 2011 and to 14.3 percent in 2021. Since interest rate spread is driven by intermediation costs and information asymmetry (Beck and Hess, 2006), an increase in the interest rate spread implies either a decrease in intermediation efficiency and/or increase in agency costs due to information asymmetry. Similarly, higher lending rates are driven by higher spreads rather than by deposit rates (World Bank, 2020) denoting the inverse relationship between bank lending rate and intermediation efficiency. On top of this, the IMF‟s financial institutions efficiency index, in which banks are a major part, indicates a declining trend particularly after 2014; from 76.5 percent in 2014 to 66.9 percent in 2021 (IMF, 2021).

Most of the studies conducted on the efficiency of the Ethiopian banking sector found an efficiency level higher than IMF results. Most For instance, Daniel et al (2023) found 94.2 percent average efficiency for the period 2014 to 2020; Ram and Mesfin (2019) found 89 percent average efficiency for the period 2010 to 2017; Dinberu and Wang (2018) found a slightly higher efficiency level, 92.69 percent, for the period 2005 to 2016. Yet others found comparatively lower level of bank efficiency. Abdurezak and Tesfaye (2016) found an efficiency level of 84 percent, using a long panel data from 1997 to 2015; Rao and Lakew (2012) found an efficiency level of 73 percent for the period 2000 to 2009. Moreover, most of the studies reviewed do not found the efficiency level that fluctuates slightly over time. Contrastingly, Emishaw (2016) found out an improvement in the efficiency level of banks over the period 2000 to 2013. None of the studies, however, indicated a decline in the efficiency level of banks. These findings seem to contrast with the increase in the interest rate spread and lending rates evinced in the banking sector over time.

All said hitherto, there still is limited research on the efficiency of commercial banks in Ethiopia. Most of the studies conducted in Ethiopia used limited observations and predominantly used non-parametric approach. Most of the studies thus far focused on measuring overall efficiency level of banks using Data Envelopment Analysis (DEA) without separating structural inefficiency from transitory inefficiency. Moreover, the effort to examine the efficiency correlates is inadequate. Finally, to my knowledge none of the studies so far attempted to measure technical progress of the banks in Ethiopia. Therefore, this study, similar to Badunenko and Kumbhakar (2017), strives to measure the efficiency level of commercial banks in Ethiopia separating persistent structural inefficiency from transitory inefficiencies, and the impact of technical progress. Besides, the study attempts to identify the correlates of inefficiency in the Ethiopian commercial banks.

1.3. Research Questions

This study will try to address the following research questions:

1. How much is the persistent and transient efficiency level of Commercial Banks in Ethiopia?
2. What factors determine the persistent and transient inefficiencies among Commercial Banks in Ethiopia?
3. How much is the level of technical progress in the commercial banks of Ethiopia?

1.4. Objective of the Study

The main objective of this paper is to measure the efficiency of commercial banks in Ethiopia and to identify determinants of commercial banks cost efficiency.

The specific objective are:

- To measure the persistent and transient cost efficiency levels os commercial banks in Ethiopia
- Measure the technical progress of the commercial banks in Ethiopia
- To examine factors which are associated with cost efficiency among commercial banks

1.5. Scope and Limitaton of the Study

There are several efficiency measures for banks such as cost efficiency, revenue efficiency, profit efficiency, technical efficiency and allocative efficiency. This study is limited by availability of data for some banks for some years. Even though, this study envisaged to measure cost efficiency of banks over the period 1996 to 2021, due to limitation of data the study covered the period from 1998 to 2021. Besides, for some of the banks the data coverage in terms of period does not match with their age. For Wegagen Bank the data covered from 2007 onwards as the annual reports before 2007 could not be obtained and those available at NBE are not complete. Similarly, for Cooperative Bank of Oromia the data covered from 2008 onwards even though the bank started operation in 2004. These missing data could affect the results.

1.6. Organization of the Study

The remaining part o the paper is organized as follows: chapter two presets the review of related literature; chapter three contains methodology; chapter four discusses the results of the study; and finally, chapter five presents conclusion and recommendation.

Chapter Two: Literature Review

2.1. Theoretical Review

2.1.1. Theoretical Foundations of Efficiency

The concept of efficiency dates back to the classical microeconomic theory. In welfare economics efficiency is prominently understood from the perspective Pareto efficiency and Kodar-Hicks. According to Pareto efficiency, also called Pareto optimality, an allocation is efficient when no individual can be made better off without making someone else worse off. It focuses on improving the welfare of individuals without decreasing the welfare of other individuals. Kodar-Hicks efficiency is an extension of Pareto efficiency, where a change is considered efficient if the winners could, in theory, compensate the losers and still be better off. Kodar-Hicks efficiency perspective focuses on maximizing total welfare or wealth, regardless of how it is distributed. Eventhough, Pareto optimality has significant implications for public theory its practicality is limited by lack of quantification of efficiency (Ten Kate, 2016).

Economic efficiency, on the other hand, referes to the optimal allocation of resources to maximize welfare or output. It is associated with, and hence, analyzed from the perspectives of welfare improvements and profit maximization that is discussed under the auspices of perfectly competitive markets. The classical microeconomic theory of efficiency focuses on minimizing costs and maximizing outputs through non-wastages for a given technology and available inputs; that is on technical efficiency. Efficiency, therefore, is driven by value creation that enables the firm possess competitive advantage over less efficient competitors. The ultimate objective of profit maximization is eventually attained. Corollary to this, efficiency is defined either as success in producing as large output as possible from a given set of inputs (Debreu, 1951); the inability to increase any one of the outputs without a reduction in atleast one other output or an increase in atleast one other input (Koopman, 1951); or the ability of the firm to achieve its output with a minimum input level (Drucker, 1963). Efficiency, thus is the effectiveness in minimizing wastage of effort, time and skill (Alber et al, 2019). These definitions of efficiency presume correct measurement of all inputs and outputs (Farell, 1957). Farell interpreted efficiency as the relation between the productiveness of a subject in relation to a model subject (Kulik, 2017). Therefor, unlike Koopman (1951) Farell‟s conceptualization of technical efficiency is relative in that it is measured by comparing the achieved relationship of outputs to inputs of a firm to the best performer. This is fundamental to the measurement of efficiency.

Although very related technical efficiency is fundamentally different from productivity in that instead of actual output the maximum output for a given level of input is used to measure efficiency (Coelli et al, 1998; Bikker and Bos, 2005; Ferrier and Lovell, 1990). Maximum output, here, refers to the optimum level of output or maximum potential output obtainable from a given input that is defined in terms of production possibilities (Ferrier and Lovell, 1990). In addition, technical efficiency is determined for a production set based on a given production function whereas only data on input and output is required to compute productivity.

Economic efficiency also encompasses allocative efficiency which relates to success in choosing an optimal set of inputs with a given set of input prices (Alber et al, 2019). Farell (1957) is the first to separate economic efficiency into technical efficiency and allocative efficiency. While technical efficiency, as explained in the aforementioned paragraph, deals with the physical relationship between inputs and outputs given ideal conditions; allocative efficiency is concerned with informing resource allocation decisions by taking into account both productive efficiency as well as Pareto efficiency. Allocative efficiency occurs when resources are distributed in such a way that the marginal cost of producing a good equals the marginal benefit to consumers, meaning that the price of a good reflects the cost of its production. In perfectly competitive markets, allocative efficiency ensures that resources are used in the most valued way.

Another common concept in efficiency is X-efficiency which has been proposed by Leibenstein (1966). Unlike technical efficiency, X-efficiency is defined under imperfect market conditions. X-inefficiency is an internal to the organization. According to Leibenstein (1966) X-efficiency is mainly caused by lack of motivation and lack of knowledge. Thus, more products can be produced without increasing inputs by improving efficiency through redesigning incentive structures and improving knowledge and learning. This is at odds with the neoclassical economists assertion of technical efficiency that once proper allocation of resources is made no further improvements in productivity can occure until technology advances. Liebnestein‟s X- efficiency challenges the accepted assumptions of neoclassical economists of maximizing behavior and rational agents, and homogenous inputs stating that individuals may not maintain maximizing objectives and inputs such as labor may differ by their motivation and cognitive capacity. Besides, the characterization of markets as competitive and thus eliminating inefficient firms is questionable as inefficiency in firms tends to exist over long periods (Mathews, 2010). This is mainly because because barriers to entry, regulatory restrictions and inertial factors. Therefore, efficiency of firms shall be measured through maximization of profits or minimization of costs taking organizational factors into consideration. Simply put, X-efficiency is non-maximizing behavior originating from lack of motivation and knowledge.

2.1.2. Measurement of Efficiency

Measurement of economic efficiency has gone through several phases and multiple methodologies have been forwarded over time. For long, efficiency was measured as the average productivity of labor (Farell, 1957). This, nevertheless, is inadequate as it leaves out other inputs of production . Debreu (1951) tried to resolve the measurement problem by developing „indices of efficiency" taking the weighted average of inputs and outputs. Farell (1957) identified two limitations of such index-based efficiency measurement. First, index-based measure of efficiency is affected by sensitivity to input variables; subjectivity in the weightings used; inability to capture intangible factors; difficulty of interpretation and issues relating to the normalization of base years. Second, this approach attempts to measure efficiency on the assumption that the efficient production function is known through technical specification of technical experts such as engineers. However, such production functions lack accuracy due to human mistakes as the production process becomes complex. In short, a theoretical function that could be used as a standard of efficiency is impractical.

In light of the aforementioned limitations and drawing inspiration from Debreu (1951), Farell (1957) pioneered a method of measuring efficiency empirically. The measurement developed by Farell (1957) takes account of all inputs Farell redefined the measurement standard of efficiency as the best achievement determined through observing inputs and outputs of firms in an industry. This brought a paradigm shift from a theoretical function based to a frontier based measurement of efficiency. The frontier based analysis is a sophisticated way of benchmarking relative performance among firms (Berger and Humphrey, 1997). This brought several benefits: it bridges the gap between theory and practice (Aigner et al, 1977); allows firms with little industry experience to select and suit the best-practice firms and permits managements with institutional experiences to objectively identify areas of best practice within the production process (Berger and Humphrey, 1997).

Currently, the measurement of efficiency consensually is frontier based. Farell (1957) divided efficiency in to technical and price efficiency. Technical efficiency is success in producing maximum outputs from a given set of inputs relative to the set of firms from which the frontier function is estimated; while price efficiency refers to success in choosing an optimal mix of inputs. Technical efficiency as applied by Farell (1957) mainly reflects the quality of inputs for measuring the quality of management separately from other factors is impossible. In contrast, Liebenstein (1966) argues that significant portion of the level of unit cost depends on internal organization of a firm and its response to external factors which he termed as X-efficiency. X- efficiency relates to the efficiency in the internal organization of the firm (such as moral, bureaucratic inertia and human error) and a firm‟s response to external factors such as competitive pressure. It indicates the ability of the management to convert inputs efficiently into outputs given prevailing prices (Bikker and Bos, 2007). Therefore, X-efficiency is a measure of managerial efficiency in executing production plans. Data suggest that gains in X-efficiency from responses to competitive pressure in the form of effort, search and utilization of new information are significant while the gain from increasing allocative efficiency is insignificant (Alber et al, 2019).

2.1.3. Performance Measurement in Banking

Explaining performance of banks is fundamental to the understanding of bank operations. The performance of banks can be explained through market power, incentive problem and inefficiency (Bikker and Bos, 2005). The market power explanation endeavors to associate bank performance with the structure of competition within the industry. The market power explanation asserts that in the absence of perfect competition, firms can maximize profit without minimizing costs; thus changes in profitability over time and between banks can be explained by the market structure among banks (Bikker and Bos, 2008).

However, even under the existence of perfect competition, banks may fail to attain profit maximization due to incentive problem or as normally called the principal agent problem. The source of principal agent problem is information asymmetry between management and capital providers of a bank. That is, shareholders and creditors of a bank may not have complete information about the operations of the bank and behavior of the management. The principal-agent theory states that shareholders are unable to adequately monitor bank management and that resulting managerial discretion may induce suboptimal behavior; in effect profits are not maximized and/or costs are not minimized (Bikker and Bos, 2008). Principal-agent problems are particularly important for banks because unlike other firms banks not only have dispersed shareholders but also dispersed creditors or deposit holders (Dewatripoint and Tirole, 1994 cited in Bikker and Bos, 2008).

Finally, failure of banks to attain maximum profit could be induced by inefficiency of the bank, the inefficiency view. This view claims that a bank earns higher profits if it makes better use of its inputs and transforms them in the cheapest possible way (Bikker and Bos, 2008). Thus, the difference in the performance of banks rests on efficiency of management in allocating and technical use of inputs, scale, and scope of operations. Under this view, efficiency plays a central role in explaining the differences in the performance of banks. Furthermore, this view can aid in identifying and measuring the sources driving bank performance.

Corollary to the above perspectives, banks use two major approaches in measuringr performance, non-structured and structured approaches. The non-structured approach employs ratio analysis; comparison of performance among banks; and the relationship of performance to investment strategies and characteristics of performance to capture various aspects of performance (Hughes and Mester, 2008). The non-structured approach, for instance, looks for evidence of agency problems by correlating performance ratios with banks governance. The major problem with the non-structured performance evaluation of banks is that there is no unifying general theory that provides basis for analyzing performance (Hughes and Mester, 2008). On the other hand, the structural approach relies on the economics of cost minimization or profit maximization where performance equation denotes a cost or profit function. The frontier based measurement of bank efficiency falls under the structural approach.

2.1.4. Efficiency Measurement in Banks

Efficiency measurement in banks employs a frontier based analysis (Berger and Humphrey, 1997). In the frontier based measurement, efficiency is measured by how close a firm is to the benchmark or best practice frontier (Berger and Humphrey, 1997). The best practice frontier is set using accounting data owing to the lack of engineering data (Berger and Humphrey, 1997) or due to potential biases in the engineering measurement of inputs and outputs (Farrel, 1957). The frontier based measurements differ in three major aspects. These are specification of a functional form for the best practice frontier; recognition and allowing of random errors; and when random errors are allowed the distributional assumptions of the random errors and inefficiency (Berger and Humphrey, 1997).

The two broad approaches under the frontier based measurement of efficiency are the parametric approach and non-parametric approach. The parametric approach determines efficiency by comparing a firm‟s performance with a pre-specified production or cost function using econometric analysis. The frontier developed under the parametric method takes noise into account. Hence, only deviation beyond this noise is attributed to inefficiency. On the other hand, the non-parametric method employs linear programing methods that does not need specification of a functional form. The non-parametric approach does not allow noise, hence, all deviations from the best practice frontier are attributed to inefficiency.

The parametric approach has three varieties namely the stochastic frontier analysis (SFA), the distribution free approach (DFA), and the thick frontier approach (TFA). All the parametric approaches impose a functional form for the best practice frontier and allow random errors in the measurement of efficiency (Berger and Humphrey, 1997). Their main difference lies in the distributional assumption used to disentangle the random error from inefficiency. SFA assumes inefficiencies to follow a less flexible asymmetric distribution, usually half-normal distribution, and random errors to follow a symmetric distribution. On the other hand, DFA allows inefficiencies to follow any distribution as long as they are non-negative; and the TFA makes no distributional assumptions on either inefficiency or random error except to assume that inefficiencies differ between the highest and lowest quartiles and that random error exists within these quartiles (Berger and Humphrey, 1997). Under SFA the border efficiency represents the most efficient points; the distance of every observation relative to this border represents the degree of inefficiency and measurement error (Battese and Coelli, 1995; Resti, 1997). Measurement errors are random that are independent and identically distributed (iid.) and follow a normal symmetric distribution, whereas the inefficiency term is half normal or truncated distribution that is independent and identically distributed (iid) across different banks (Battese and Coelli, 1995; Resti, 1997).

The non-parametric approach uses a mathematical approach to measuring efficiency. It includes the Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH). DEA employs a linear programming technique that provides a linear piecewise frontier by enveloping the observed data points yielding a convex production possibilities set without an explicit specification of a functional form (Athanasoglou et al, 2008). It is an optimization technique for the evaluation of multiple-input/multiple-output firms (Rest, 1997). The FDH technique is a special category of DEA, where, instead of convexity, free disposability of inputs and outputs is assumed (Athanasoglou et al, 2008). FDH produces larger efficiency measures than the DEA technique.

The criticism of the parametric approach is the need to specify a functional form that presupposes the shape of the frontier. Such specification could confound the measured efficiency with specification error (Berger and Humphrey, 1997). The major pitfall of the non-parametric techniques is that they do not allow random error that takes measurement error such asa luck and accounting inaccuracies into account. Any deviation from the frontier is evaluated as inefficiency (Berger and Humphrey, 1997). As such noise can disturb the shaping and positioning of frontiers (Athanasoglou et al, 2008). That in a way implies errors in any one of the decision unit"s efficient frontier could alter the measured efficiency of all the units (Berger and Humphrey, 1997). The parametric approach is better choice when measurement errors and random events are significant, whereas the non-parametric approach is more appropriate when random disturbances are less of an issue and price information is unavailable (Coelli et al., 1999).

Research that compare efficiency measures under the parametric and non-parametric approaches brought disparate results. Ferrier and Lovell (1990) have found a positive but statistically insignificant correlation between technical efficiency rankings of banks using the two approaches. Resti (1997) does not find differences between the two approaches when they are based on the same data and conceptual framework. In contrast, Huang and Wang (2002) found contradictory ranking of bank scores based on the parametric and non-parametric approaches. Besides, they found out that parametric methods are more strongly correlated with the conventional ratio analysis methods of bank performance than the non-parametric approach. In general, the empirical findings under the two methods differ mainly due to difference in the structure and implementation of the approaches.

Performance measurement of banks is crucial to assess their operational efficiency and effectiveness in resource utilization. As such pertinent performance measures need to be applied when evaluating the performance of banks. As banks are profit-driven entities a measurement that assesse profit-maximization or cost minimization provides a germane basis to evaluate performance. Thus, the structured frontier based approach measurement provides a germane measurement approach for Ethiopian banks. In order to account for measurement errors and randomness of certain events the parametric approach is more suitable.

Among the parametric measures of efficiency SFA provides distnict advantages over other approaces. First, SFA allows incorporation of random error in the production or cost functions which allows more accurate estimation of the frontier. Second, SFA allows separation of inefficiency from the random error that guides decision-makers valuable insights into where improvements can be made. Third, compared to the distribution free approach SFA provides more flexibility in specifying the functional form of the production or cost frontier. Fourth, SFA allows the inclusion of variables that may influence efficiency, providing more complete understanding of efficiency levels and drivers. This facilitates targeted intervention and policy recommendations. Finally, SFA allows generation of efficiency scores for individual firms, facilitating comparison among firms in the industry. These features make SFA the preferred approach to measure efficiency of firms with complex production systems such as banks.

2.2. Empirical Literature

After, the development of efficiency measurement methodology by Farell (1957), efficiency researches have been conducted in many sectors. Banking is not an exception; banking efficiency studies have been started in early 1960s. Bank efficiency researchers have proliferated in the 1980‟s driven by increase in competition following deregulation of the banking sector in many countries, the ensuing reduction in bank profitability, and subsequent crisis in some banks. Till the present time, the dynamics and differences in the development level of banks across the world have made banking efficiency studies to remain appealing research area. The methodologies followed and the results obtained are manifold. The research concerns of studies can be categorized into determining the efficiency level of banks; comparing efficiency levels and rankings under the parametric and non-parametric approaches; comparing the relative importance of scale, scope and technical efficiency; comparing results under different functional forms and identifying determinants of bank efficiency. This section summarizes the findings of these studies.

Scores of researches that measure the cost efficiency levels of banks is conducted in the developed markets of US and Europe. Particularly, the US banking sector is studied intensively since the early 1990‟s. The abundant research in the US has established technical inefficiency of around 20% for US banks (Bikker, 2005; Bikker and Bos, 2005). Nonetheless, there are still studies in the US that significantly deviate from this result. Ferrier and Lovell (1990) employing the production approach on a sample of 575 financial institutions that participated in the US Federal Reserve Systems‟ Functional Cost Analysis found a 26% cost inefficiency. Mester (1996) after accounting for quality of assets and the risk preference of banks found 6-9% level of technical efficiency among Third Division Banks in the US. Evanoff and Ors (2008) found inefficiency level of 14.4% among non-merging „Incumbent Banks‟ in the US. Mester (1996) after taking risk preferences of banks and quality of assets found 6-9% inefficiency level for the US banks.

Several studies have been conducted in Europe too. In Italy, Resti (1997) and Girardone et al (2004) have found technical inefficiency level of 30.2% and 13-15% respectively. Lang and Welzel (1996) have evaluated the inefficiency and technical progress of cooperative banks in Germany. They classified the banks into 10 categories based on size of the banks. The result indicates that the most efficient class has 87% efficiency while the least efficient has 57% efficiency. Athanasoglou et al. (2008) has measured the inefficiency level in Greek. In Asia studies by; Huang and Wang (2002) in Taiwan; McKillop et al (1995) and Atlunbas et al. (2000) in Japan; Badunenko and Kumbhakar (2017) in India have evaluated the efficiency level of banks in India

International comparison of bank performances is difficult and has several intricacies due to differences in the accounting standards and systems, regulation, and differences in banking models. Notwithstanding this, cross country comparison of bank inefficiencies has also been attempted. The results international studies are diffcut to generalize. In some studies the inefficiency results are close to the single country studies while in others it is uncomparable to single country studies. Atlunbas et al. (2001) obtained technical inefficiency of 20-25% for the European market and Maudos et al (2002) found inefficiency levels of 17.5% for European banks. Allen and Rai (1996) found higher recorded efficiency levels for countries that allow banks to engage in non-traditional banking than those countries that do not allow. The earlier has an inefficiency of 15.1% while the later has an inefficiency level of 21.1% . In contrast, Bikker (2005) found an inefficiency level significantly far from the aforementioned studies. In his study, Bikker (2004) found as high inefficiency as 61.9% when using single frontier for the Europe area. When they account for country-specific frontiers and bank specialization the inefficiency improves to 50.7% and 37.8%, respectively. In general, both with single country studies and cross-country studies the inefficiency results are fairly comparable; in most cases the results fall between 15 and 30%. There are exceptions to these results though.

Scale and scope efficiencies are the concern of many empirical studies. The focus in scale efficiency studies is its existence and/or relative importance of scale efficiencies with technical efficiencies. Bikker and Bos (2005) assert that X-efficiency is quantitatively more important in the banking industry. Atlunbas et al (2001) and Lang and Welzel (1996); and Mester (1996) have shown that technical efficiency is far more important than scale inefficiencies in determining the overall efficiency of banks. The scale inefficiency among European banks is between 5 and 9% while the technical inefficiency amounts to 25 to 30% (Atlunbas et al., 2001). Berger et al (1993); Mitchel and Onvural (1996) have similar conclusions that X-efficiencies brought about by management technologies and other factors exceeds those efficiencies resulting from scale and scope economies. Berger et al (1993) found that X-inefficiencies account for 20% or more of the costs in banking while scale and scope economies account for less than 5% of costs. Berger and Humphrey (1997) and Lovell (1993) also argued that X-efficiency is important because it bears the advantage of allowing the bank to react to price changes. Mester (1996) did not find evidence of scale and scope economies among Third Division banks in the US that implies no cost efficiency gains by changing the loan mix of banks or by changing its level of output.

In addition to the evaluation of inefficiency levels, several studies have attempted to identify the determinants of bank inefficiency. These studies regress bank level characteristics and environmental variables with the inefficiency levels of banks. The commonly used variables include bank size, branch network, funding risk, bank ownership and profitability. The case of nonperforming loans and equity is quite unique. In some researches they are considered as part of the function used to determine the efficiency frontiers and in others they are employed as inefficiency correlates. In addition, variables such as asset concentration as measured by intensity of investing in government debts or disintermediation (For example, Mosunda, 2008) and geographic location (Girardone et al., 2004) are used. Macro level variables such as market characteristics, intermediation ratio and regulation intensity, inflation (Mosunda, 2008) and contribution of the banking sector to GDP, measured as the ratio of deposits and credits of banks to GDP, (Badunenko and Kumbhakar, 2017) are employed. Mosunda (2008) found a positive association between credit risk, increase in interest bearing deposits, disintermediation or bank debt to the government and macroeconomic instability, and inefficiency. On the other hand, ownership, opportunity cost and regulatory pressure do not have significant impacts on inefficiency. Girardone et al. (2004) report no clear relationship between asset size and bank inefficiency. Very recently Badunenko and Kumbhakar (2017) conducted a study to identify inefficiency correlates of Indian banks. In particular, they have assessed bank specific factors such as ownership and bank regulations.

2.2.1. Sub Saharan Africa

The 1980s and 1990s is characterized by banking reforms and financial liberalization and the ensuing increase in bank failure in many parts of the world. Intensive research on banking followed this era of liberalization all over the world, particularly in western countries and Japan. Contemporarily, a number of African countries have began to restructure their financial sector in order to boost banking efficiency (Brownbridge and Harvey, 1998 cited in Kiyota, 2011). In spite of the intense research in other parts of the world, bank efficiency studies in Sub-Saharan Africa (SSA) is evidently scanty (Tefula, 2001; Mosunda, 2008). In a comprehensive survey of bank efficiency researches by Berger and Humphrey (1997), out of the 130 bank efficiency studies 124 are industrialized countries and only six studies are from developing countries; none are from Africa. With such dearth of studies, there is limited evidence to generalize about the efficiency of banks in SSA.

The important issues that stand out in bank efficiency studies of SSA are ownership, including foreign banks entry, banking reforms and market structure, and bank size. The studies conducted in SSA have accorded particular significance to foreign banks and efficiency, specifically on the association of foreign bank entry and ownership, with efficiency. Studies by Kirkpatrick et al, 2008; Kiyota, 2011; Mwega, 2011 and Haunters and Pieris, 2021 explored this particular issue. Foreign bank entry, level of foreign banks penetration in a country, level of foreign ownership in a bank and origin of the foreign bank are some of the factors assessed in these studies. Comparison of cost efficiency between domestic and foreign banks is also quite common. Moreover, the impact of competition and other bank level characteristics are studied. Capitalization, bank size, management of funds and asset quality are among the commonly studied bank level characteristics.

In comparison, foreign banks are more efficient than domestic banks in SSA (Kirkpatrick et al, 2008; Mwege, 2011). Not just more efficient but the degree of foreign bank penetration improves cost efficiency of banks in SSA countries mainly due to the strengthening of the regulatory system and increased scrutiny of banks when foreign bank ownership is substantial (Kirkpatrick et al, 2008). With regard to the origin of foreign banks, Sub-Sahara based foreign banks are more efficient than non-Sub-Sahara based foreign banks (Kiyota, 2011). This underscores the importance of cultural and geographic closeness to the performance of foreign banks.

In relation to efficiency correlates; level and adequacy of financial capital, bank size, management of funds, and market capitalization are among the fundamental bank level characteristics studied in the context of SSA. The findings on size of a bank are inconclusive. Negative relationship between bank size and bank inefficiency is found smaller banks are more efficient than larger banks (Akoena et al, 2009; Ncube, 2011; Kiyota, 2011). Ncube (2011) ascribed this finding to the lower overhead cost, provision of fewer basic services and use of fewer personnel by smaller banks. Conversely, in a cross country study of SSA countries Kirkpatrick et al., (2008) found out higher efficiency for larger banks. Similarly, Ally and Patel (2014) and Haunter and Pieris (2021) have found higher efficiency for larger banks in a Tanzania and Uganda, respectively. The higher efficiency of larger banks emanates from the capacity of larger banks to attract more quality employees (Kirkpatrick et al., 2008). The effect of financial capital on efficiency is contradictory in SSA. Capital to total assets ratio is found to have a negative impact on the efficiency of SSA banks (Kablan, 2011) and Kenyan banks (Ikapel et al, 2023). A negative impact on bank efficiency could be related to the moral hazard agency problem (Kablan, 2010). Level of capitalization measured by book value of capital to total assets is also found to have a negative association with bank efficiency (Kablan, 2010; Ikapel et al, 2023). On the other hand, Kiyota (2011) found positive association between financial capital to total assets ratio and bank efficiency on foreign banks operating in SSA. The result for bank leverage measured by the ratio of equity to total assets has mixed results for domestic, SSA based and non SSA foreign banks. It has a negative relation with cost efficiency for domestic banks and non-SSA foreign banks while it affects SSA foreign banks positively (Kiyota, 2011). Further, Kiyota (2011) has found management‟s use of funds measured by the ratio of fixed assets to total assets does not show significant association with bank efficiency in SSA countries.

The findings on bank efficiency in East Africa are contrasting. The efficiency of banks has evinced improvement over the years in Kenya (Mwega, 2011; Samuel and Kamau, 2021) and in Tanzania (Ally and Patel, 2014) in Tanzania. In contrast, using interest rate margin and spreads Cihak and Podpiera (2005) concluded that the efficiency of East African banks does not improve despite reforms and the presence of international foreign banks. Employing similar methodology Beck and Hesse (2006) found no improvement in efficiencies after banking sector in Uganda.

2.2.2. Banking in Ethiopia: history and system

The history of banking in Ethiopia has several turns and twists. Since its beginning, banking in Ethiopia has passed through changes in ownership, laws, and systems. The first modern bank in Ethiopia, the Bank of Abyssinia, had been established in 1905 by the Anglo-Egyptian National Bank (Belay 1987 cited in Muluneh 2006; Alemayehu, 2006). The bank has been given the right to issue notes and coins. Two additional banks were also established within a decade of the establishment of the Bank of Abyssinia. Paradoxical to the current situation in Ethiopia, all these three banks, by then, were foreign owned (Alemayehu, 2006). Although Bank of Abyssinia has been given a monopoly right for fifty years, it was wholly purchased by the government of Ethiopia and renamed the Bank of Ethiopia in 1931(Gidey, 1990 cited in Alemayehu, 2006 and

Muluneh, 2006). However, Bank of Ethiopia was short-lived due to the Italian occupation. During the Italian occupation, banking activity expanded (Alemayehu, 2006). But this expansion was temporary because following withdrawal of the Italians from Ethiopia, a new bank, the State Bank of Ethiopia, was established in 1942 and started operation in 1943 (Muluneh, 2006). The State Bank of Ethiopia performed the functions of both a central bank and commercial bank until 1963 (Alemayehu, 2006).

In 1963 a new banking law was enacted. This law split the functions of the State Bank of Ethiopia into central banking and commercial banking and allowed other commercial banks to operate. Further, the 1963 law allowed foreign ownership of banks up to 49% (Belay, 1987 cited in Muluneh 2006). Nevertheless, this period of liberalization has not sustained due to the nationalization of all the private banks following the downfall of the Imperial Regime and the ensuing establishment of a socialist government. During the period of state socialism the main aim of the financial institutions was to execute the economic plan of the government and direct resources to the public sector (Alemayehu, 2006; Muluneh, 2006).

Finally, after the fall of the socialist regime, the bank sector is reformed under Proclamation No. 84/1994. Corollary to this reform several private banks have been established. Since the banking sector has recorded an immense growth in terms of resource mobilization, loan disbursement, capital, and outreach or branch network. By 2022 the number of banks in Ethiopia has reached 30, two state and 28 private banks, with aggregate capital of Birr 199 billion and total bank branches of 10,221, making the population to bank ratio of 10,289.2 (NBE, 2022).

The banking reform in Ethiopia followed a gradual strategy (Alemayehu, 2006). This contrasts with the reforms implemented in most African countries which followed more radical reforms (Brownbridge and Harvey, 1998 cited in Muluneh 2006). The 1994 reform has not allowed foreigners to own banks; full liberalization is postponed until the domestic capacity is strengthened (Alemayehu, 2006). This has made the Ethiopian banking system unique from other SSA countries (Kiyota, 2007). The banking sector in Ethiopia is characterized by a dichotomy of private and state ownership. Unlike Ethiopia, the banking sector in other parts of SSA is characterized by the presence of foreign banks. This is due to the long-held prohibition of foreign banks to operate in the country. However, very recently, the government of Ethiopia has started to relax the operation of foreign banks and with Proclamation 1159/2019 lifted the legal restrictions in place on ownership of bank shares by foreign nationals of Ethiopian origin.

For a long time, the Ethiopian banking sector has been dominated by the State-owned banks which mobilize significant portion of total assets, deposits, loans and capital. For instance, by 2012 the share of private banks from the total deposits liability was only 32.3%; state-owned banks account for the remaining 67.7% (NBE, 2017). By 2017 private banks account for 34.4% of the total deposit liability and state-owned banks account for the remaining 65.6% of the total deposit liability (NBE, 2017).Similarly, as Kiyota (2007) noted the share of private banks from the total assets of the banking sector by 2006 is only 30.4%.

The dominance of state-owned banks, however, is recently waning. For instance, by 2022, the share of private banks has surpassed state-owned banks in terms of branch network, capital, and fresh loans disbursement. The private banks account for 62.9% of the fresh loans disbursed during 2022. In 2022, private banks account for 73.1% of the total bank branch network in the country compared to 23.9% for state owned banks. With regard to total capital of the banking sector, private banks account for 58.5% of the total capital. Until 2013, the state-owned banks had more branch network (51.4%) than private banks. In other yardsticks such as deposit mobilization and outstanding credit the gap between state-owned banks and private banks is narrowing rapidly. Private Banks account for 47.7% and 46.7% of total deposit liabilities and outstanding credit in 2022 (NBE, 2022).

2.2.3. Efficiency of Ethiopian Banks

Everywhere else in the world, financial liberalizations and banking sector reforms have been followed by extensive research on efficiency mainly because efficiency is one of the targets sought. That is not the case in Ethiopia; very few research have been conducted on the efficiency of Ethiopian banks. Even, the few conducted came long after the financial reform in the country. The efficiency studies try to accomplish two basic objectives: measuring the level of efficiency of banks and examining the factors that determine efficiency among Ethiopian banks.

A review of extant literature on the efficiency of Ethiopian banks reveals some important features. First, methodology wise, most of the studies conducted employed the Data Envelopment Analysis (DEA) approach. Rao and Lakew, 2012; Abdurezak and Tesfaye, 2017;

Ram and Mesfin, 2019 and Dinberu and Wang, 2018 have used DEA approach. Conversely, the stochastic frontier analysis is uncommon. Second, the results in terms of efficiency measure are inconsistent with high variation in the efficiency measures of banks among studies. Third, there is high interbank variation in the efficiency level of banks. Fourth, there is inconsistency in the efficiency comparison of private and state-owned banks.

In terms of efficiency level the results range between 73% and 92.7%. Using DEA approach Rao and Lakew (2012) have found average cost efficiency of 73% for the period 2000 to 2009. On the other hand, Dinberu and Wang (2018) found an average efficiency level of 92.7% under the variable returns to scale assumption and 80.6% under the constant returns to scale assumption for the years 2005 to 2015. This result is closer to the finding by Tesfaye and Abdurezak (2017) using a long panel data from 1999 to 2015 found an average efficiency of 84% and 92% under the constant returns and variable returns to scale, respectively. Ram and Enyew (2019) using a DEA approach have found an overall efficiency level of 89% for the period 2010 to 2017. In contrast, the few studies that have used stochastic frontier analysis resulted in comparable levels of efficiency. Muluneh (2006) using a stochastic frontier analysis for the period 1998 to 2006 found an average efficiency score of 90.8% while Emishaw (2016) using panel data for the period 2000 to 2013 found an efficiency score of 92.7%.

Probable reasons for this inconsistency could be the difference in the periods covered by the studies and the variation in the measurement of inputs. Particularly, the measurement of labor price varies among the studies. In some of the studies such as Abdurezak and Tesfaye (2017) and Tadele (2016) the price of labor is left out altogether. Yet others have measured the price of labor as the ratio of salary and benefits to total assets such as Rao and Lakew, 2012; Dinberu and Wang, 2018. Such variation in the measurement of labor price is uncommon in bank efficiency literature worldwide. Further, distinct from other studies, Rao and Lakew (2012) has taken deposits as an output rather than as an input that could have caused the difference in measured efficiency.

On the other hand, high interbank variation is observed in the efficiency scores of banks (Dinberu and Wang, 2018; Ram and Enyew, 2019). The efficiency score of banks in the study by Ram and Enyew (2019) range between 100% and 11%. In addition the cost efficiency among the

Ethiopian Banks has a fluctuating trend (Dinberu and Wang, 2018; Rao and Lakew, 2012). Abdurezak and Tesfaye (2017) found the fluctuation in efficiency only among private banks. Besides, there is no improvement in the efficiency of Ethiopian banks over time (Abdurezak and Tesfaye, 2017). Based on size, the efficiency of small private banks is growing over time more than medium private banks (Abdurezak and Tesfaye; 2017).

Finally, comparison of efficiency between private and state-owned banks provides contrasting results. In some cases, state-owned banks are found to have higher efficiency than private banks (Abdurezak and Tesfaye, 2017). Abdurezak and Tesfaye (2017) has also found out that there is no convergence between private and state-owned banks over time. However, in another study Rao and Lakew (2012) found out higher efficiency scores for private banks than the state-owned banks. This is in line with Kiyota‟s (2007) finding that private banks have better performance than government banks for the period 1998 – 2006 as measured by return on asset (ROA) and interest rate spreads.

The common bank level characteristics assessed to examine the determinants of bank efficiency include size of a bank, profitability, deposit, capital adequacy, and loan quality. In addition, age (Muluneh, 2006); ownership, market share and market concentration (Rao and Lakew, 2012); diversification (Tadele, 2016) and intermediation ratio (Emishaw, 2016) are also examined. The results for the relationship between size and bank efficiency are inconsistent. The studies in relation to size of bank have found positive association (Rao and Lakew, 2012) and Muluneh, 2006); others found insignificant association (Tadele, 2016 and Emishaw, 2016) and yet others found negative relationship (Abdurezak and Tesfaye, 2017). Similarly, the results for financial capital are inconsistent. Rao and Lakew (2012) and Muluneh (2006) found positive relationship between bank efficiency and capital, while Emishaw (2016) found negative association. Profitability as measured by ROA relate positively with bank efficiency (Rao and Lakew, 2012; Emishaw, 2016). Abdurezak and Tesfaye (2017) found out that deposit growth rate, loan size and growth of earning assets negatively affect bank efficiency. In their research they found out that banks holding excessive deposit are disadvantaged in terms of efficiency. Tadele (2016) found no significant association between loan quality and diversification with bank efficiency.

In general, extant literature indicates that, with few exceptions, the cost inefficiency level of banks range between 15% and 30%; whereas scope and scale inefficiency levels are between 5%

to 10%. These findings evidence the importance of x-efficiencies compared to scale and scope efficiencies. Consequently in Ethiopia meaningful contributions can be made by prioritizing the cost efficiency of banks. The studies conducted in SSA emphasize, and take into account, the importance of foreign ownership oin bank efficiecncy. Nevertheless, until 2023 neither foreign banks nor ownership of banks by foreign nationals is allowed in Ethiopia. The context of banking in Ethiopia is disparate to that of SSA countries. Hence, unlike SSA countries efficiency study in Ethiopia, at least for the time being, need not be to account for influence of foreign banks penetration.

Previous bank efficiency studies in Ethiopia measure labor price as the ratio of salary to total assets. Some studies have even left labor price altogether. Such measurements do not properly reflect the actual price of labor and could bias the results. This study opts to resolve this limitation by measuring labor price as the ratio of salary and benefits expense to the number of employees. This is a measurement used in the majority of researches. Finally, several studies nave employed bank specific characteristics such as size, capital level, age and intermediation to identify efficiency determinats. In this study, too, bank specific characteristics are employed to assess efficiency correlates of banks.

Chapter Three：Methodology

This chapter discusses the major methodological issues in banking efficiency. The overall methodological framework for the stochastic frontiers analysis is determined by the general method selected (parametric or non-parametric); the inputs and outputs of banks used; the data employed (cross-section or panel data) and the specification of the functional form and the assumption regarding the correlation of firm effects with regressors. This section presents these issues one by one and forwards the method specified for this study.

3.1. Specification of Inputs and Outputs of Banks

The measurement of efficiency in banks, like any other organization, depends on comparison of inputs and outputs. The bank efficiency literature is in disagreement on the definition of inputs and outputs and what constitutes inputs and outputs of banking. Benston (1972) identifies absence of readily identifiable products, difference in the research purposes, and availability of data as the major reasons for the disagreement. Unlike the case in manufacturing industries, neither the production process nor the outputs of banks are readily discernible. In addition, the purposes for which bank efficiency studies are conducted could call for differences in the definition of outputs in banking. Some are conducted to assess the internal management and other to assess overall competitiveness of banks. Such differences could call for different input- output definitions. When assessing overall competitiveness and viability, loans and earning assets are used as an output while when assessing the cost effectiveness of a bank number of documents processed or physical outputs are used as an output (Humphrey, 1997). Generally, the approaches on the definition of bank inputs and outputs are classified into the production approach and the intermediation approach (Humphrey, 1997; Bikker and Bos, 2005).

The production approach views a bank‟s operation as any physical production processes that takes various inputs and delivers physical outputs. The production approach distinguishes labor and capital as an input and number of processed bank documents as an output (Biker and Bos, 2005). The basis for the production approach is Benston‟s (1972) definition of bank outputs. Benston (1972) (and later Benston et al., 1982) defined outputs as what banks do that causes them to incur operating costs. Banking outputs are, thus, defined in relation to the factors that cause banks incur operating costs such as document handling, bookkeeping costs, customer handling costs. Under this approach, as in manufacturing processes, inputs and outputs of a bank are fundamentally conceptualized as flow variables such as the number of checks cashed, the number of payments made on saving deposits and the number of loans issued (Resti, 1997).

In practice, flow variables are unobservable and data is difficult to obtain. Benston (1972) suggested simple average of the number of accounts outstanding at the end of each month. Alternatively, output stocks such as the number or amount of deposits and loans are employed in place of output flows with the assumption that unobserved output flows are proportional to the flow of inputs (Humphrey, 1997). Monetary flows, the interest flow taken from profit and loss statements, can also be employed as proxy for physical flows. However, using monetary flows may bias the output measure as the interest rate associated with deposits and loans could differ among banks (Resti, 1997). Utilization of output stocks as a proxy for the physical flows is also justified on the basis that these stocks require the production of payment, provision of liquidity and monitoring of credit decisions. Further, using stock of deposits and loans instead of flow of interests associated with them avoids the bias due to different rate at which these products are sold by the banks (Resti, 1997).

The intermediation approach views banks as financial institutions that intermediate funds between fund surplus economic agents and fund deficit economic units (Sealey and Lindley, 1977; Berger and Humphrey, 1997). The intermediation approach aligns to the traditional core functions of financial institutions (Biker and Bos, 2005). Sealey and Lindley (1977) defined outputs in relation to technical and economic production processes. Accordingly, technical production is related to a transformation where inputs lose their identity to generate outputs while economic production is related to the creation of a product that is highly valued than the original inputs (Sealey and Lindley, 1977). Specifically, only products that are highly valued in the market than their original inputs are economic outputs. Earning assets are normally considered more valuable than the deposits used to produce them; therefore, only the services associated with earning assets are considered outputs of a bank.

The intermediation approach takes labor, material, capital, and deposits as inputs and earning assets as an output (Sealey and Lindley, 1977). The use of dollar values to measure physical or service flow is one major criticisms of this approach. Sealey and Lindley (1977) disagree with the idea of viewing balance sheet amounts of earning assets as stock variables. They, citing

Pesek (1970), argue that bank money is constantly sliding into non-existence as it is returned to currencies or as loans expire. In addition, continuous activity is required to ensure the continuing existence of earning assets and deposits to prevent the stock or flow of these entries from being destroyed. In fact, a financial institution can maintain a stock of earning assets and deposits on its balance sheet only by constantly incurring a flow of costs and through continuous effort. Finally, unlike the stocks known in manufacturing firms, earning assets in banks earn income. Through an empirical analysis Hughes and Mester (1993) have concluded that deposits are outputs.

The major difference between the production approach and the intermediation approach is the treatment of deposits. The production approach treats deposits as outputs of banks for two basic reasons. First, deposits are creation of values for which customers bear the opportunity cost and/or at times pay fees and charges for it. Second, it is not reasonable to exclude deposits that cause significant portion of banks' operating costs from inputs (Resti, 1997). In contrast the intermediation approach takes deposits as economic inputs on the justification that deposits are used as an input to produce earning assets, and the bank incurs costs on them without earning any direct revenue. Sealey and Lindley (1977) further argue that the safe keeping, check clearing and bookkeeping services provided to depositors are partial payments for the use of funds.

Pragmatically, however, the choice between the production approach and the intermediation approach depends on the issue the researcher is trying to address (Ferrier and Lovell, 1990). The production approach is appropriate for bank decision units or branches with low autonomy in revenue generation or loan policy (Ferrier and Lovell, 1990). The production approach is more appropriate when evaluating the cost-efficiency of decision making units while the intermediation approach is employed to address questions of economic viability.

With regard to inputs and outputs of a bank, this study adopts the intermediation approach that takes loans and earning assets as an output and labor, physical capital, and deposits as inputs of the banking process. The outcome used in the study is total cost of banks that incorporates personnel cost, operating costs, and interest expenses taken from the income statement.

3.2. Basic Stochastic Frontier Model

The basic stochastic frontier analysis model is concurrently proposed by Aigner et al (1977) and Meeusen and van den Broeck (1977). This model extends the model previously forwarded by Aigner and Chu (1968) that lacks the stochastic element. Since its initial proposal, considerable research has been conducted applying the stochastic frontiers analysis model. Unlike the previously developed deterministic models, the stochastic frontier analysis model employs a composite error term. The model is presented as follows:

$$Y_i = (Xi;\, \beta) + \varepsilon_i \qquad i= 1,2,\ldots N \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

X_i is a vector of inputs; β refers to a vector of unknown parameters to be estimated and ε_i is a composite error term less than or equal to zero. The composite error term has two components v_i and u_i such that:

$$\varepsilon_i = V_i + U_i \quad i= 1,2,\ldots N \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

v_i is the random component of the error term that represents the symmetric disturbance or noise that is assumed to be identically, and independently distributed as $N(0, \sigma_v^2)$. This random component generally accounts for measurement error, misspecification, and the randomness of the production process. u_i is an error component that represents the measure of inefficiency distributed independently of v_i. The inefficiency component, u_i, aims to represent the reduction in observed output of a firm from the maximum feasible level or the frontier production. Note that, the output level achieves maximum efficiency whenever u_i is zero; that is, the production function for the maximal output attainable by a production process given a set of inputs and the level of technology is $y_i = f(x_i; \beta) + v_i$. This frontier production function, unlike the deterministic function employed previously by Aigner and Chu (1968), is clearly stochastic.

3.2.1. Duality Theory

The stochastic frontier analysis based on the production frontier function relates outputs with inputs of a firm. Nonetheless, banks provide services that is not easily observable. Hence, bank efficiency cannot be analyzed using the production frontier function. Instead, cost frontier functions are employed. The duality condition between production and cost functions ensures that cost functions can contain the same information about the production possibilities (Shephard, 1970). Under exogenous prices and optimal production, the properties of the production function can be studied indirectly, through cost functions. Thus, stochastic frontier analysis model based on the cost function that estimates the minimum cost frontier can be used to evaluate efficiency of banks (Resti, 1997). Similar to the model based on the production frontier function, the stochastic frontier model based on the cost function will have a composite error term subdivided into two: a component representing inefficiency and a term that accounts for noise. The stochastic frontier model using cost function, thus, run as:

$$C_i = C(y_i\ ;\ w_i) + \varepsilon_i = C(y_i\ ;\ w_i) + v_i + u_i \qquad (3)$$

c_i is the cost of a bank; y_i is the bank outputs; w_i is the price of inputs; v_i is the uncontrollable random disturbance or noise and u_i is the inefficiency measurement.

3.3. Panel Data Analysis

The basic model by Aigner et al. (1977) applies for a cross-sectional data. Schmidt and Sickles (1984) identify three problems of using a cross-sectional data for stochastic frontier analysis. First, cross-sectional stochastic model lacks consistency as an estimator of individual efficiency. The whole error term which contains both the noise and inefficiency can be estimated consistently; however, the variance of the distribution of the inefficiency, conditional on the whole error term, does not vanish even when the sample size approaches infinity. Second, estimation of the model and separation of the inefficiency from statistical noise requires specific distributional assumptions for the noise and inefficiency. It is not clear how robust the results are to these distributional assumptions. Third, the model assumes the inefficiency term is independent of the model regressors. This assumption is unlikely because firms make adjustments as on input choices, if the inefficiency is known to them.

Panel data structures are helpful in resolving the above drawbacks. Panel data offers the advantage of taking into account some of the potential heterogeneity that cannot be controlled using cross sectional models. Adding more observations on the same firm yields information not attainable by adding more firms; essentially, the constancy of the inefficiency over time can be considered as evidence of inefficiency (Schmidt and Sickles, 1984). Schmidt and Sickles (1984) further assert that a panel data do not need a strong distributional assumption as are the cross-section data. Panel data provides the possibility of estimating parameters without assuming that inefficiency is uncorrelated with regressors. Besides, information on firms over time enables one to examine whether inefficiency has been persistent over time or whether it is time-varying (Kumbhakar, et al., 2014). Finally, panel data provides a more realistic characterization of the inefficiencies (Belotti et al., 2013). Consequently, much of the extant literature on stochastic frontier analysis on bank efficiency employs a panel data structure.

The first effort to extend the stochastic frontier analysis to a panel data model has been made by Pitt and Lee (1981) and Schmidt and Sickles (1984). Both have suggested a time invariant stochastic frontier analysis model using a panel data set. The general model runs as:

$$Y_i = (Xit;\ \beta) + v_{it} + u_i \qquad i = 1, 2, \ldots, N \text{ and } t = 1,2, \ldots T \text{.........................(4)}$$

Pitt and Lee (1981) proposed maximum likelihood estimation, with a truncated normal and normal distribution for inefficiencies and noise respectively. The distribution assumption for the inefficiency for u_i is crucial; without such assumption it would be difficult to separate it from the intercept. The model provides a time invariant inefficiency. The model by Pitt and Lee (1981) does not fully solve problem of correlation between the inefficiency and inputs under the cross-sectional stochastic frontier model. As the inefficiency stays with the firm overtime, there is a possibility that firms would learn about it and adjust their input choices creating a correlation between the inefficiency, u_i, and the inputs used in the model.

As a solution, Schmidt and Sickles (1984) proposed ordinary least squares (OLS) that employs standard estimation methods under the fixed effects and random effects models to relax the imposed assumptions under cross sectional stochastic frontier analysis models. Under the fixed effects, the model can be estimated using (OLS) technique after creating the necessary dummies that control for fixed effects. The advantage of the fixed effects model is to allow the inefficiency to freely correlate with the regressors (Schmidt and Sickles, 1984). This could be desirable for empirical studies in which inputs are believed to be correlated with the inefficiency (Kumbhakar et al, 2015). The major drawback of this model is that it does not allow time invariant regressors.

Generally, the methods proposed by Pitt and Lee (1981) and Schmidt and Sickles (1984) will not enable us to estimate time-varying inefficiency measures. This is unreasonable especially when the time period of the panel is long. Under this method the inefficiency level only differs across individual firms but remains constant across time. A constant inefficiency over time implies an inefficient firm never learns over time (Kumbhakar et al, 2015). Particularly, it is implausible for a firm‟s efficiency to stay constant for an extended period. Thus, there is a need to consider models that accommodate change in inefficiency over time.

Consequently, researchers proposed time dependence model that begets time-varying inefficiency. These models primarily tried to address the time invariance limitation of the Pitt and Lee (1981) and Schmidt and Sickles (1984) models. By extending the Schmidt and Sickles (1984) model, Cornwell et al (1990) proposed an individualized slope parameter stochastic frontier model in which model parameters are estimated by extending the conventional fixed effects and random effects panel data estimators. In this model, the individualized intercepts are modeled as a function of time where the coefficients are firm specific. The parameters of this function are estimated by regressing residuals from the general model $Y_{it} = \beta_0 + x'_{it} \beta + v_{it} - u_{it}$ i= 1,2,….N and t = 1,2,…, T for each unit on a constant, time and time squared; then the fitted line from this model provides a consistent estimate for the individualized firm-specific intercepts. The efficiencies for individual firms for each period are estimated using an analogous procedure as Schmidt and Sickles (1984). The quadratic specification allows for a unit-specific temporal pattern of inefficiency (Belotti, 2013). Basically, the Cornwell et al (1990) model measures efficiency by comparing a firm‟s efficiency with the best performer for that year. The best performer firms could be different in different years. However, if the maximum is defined over all firms and time; the comparison is with the best firm in the sample defined over all periods (Kumbhakar et al, 2015). The Cornwell et al (1990) model has some problems. First, the model requires estimation of a large number of parameters (Belotti et al., 2013). Second, as the model relies on standard panel data methods, the time variable that appears in the inefficiency function cannot be included as a regressor. Thus, the model cannot separate technological change from technical efficiency (Kumbhakar et al, 2015).

Kumbhakar (1990), and Battese and Coelli (1992) proposed a time varying model, also called the time decay model, that is popular in stochastic frontier analysis that employs a panel structure.

They extend the model proposed by Pitt and Lee (1981) by allowing the mean of inefficiency to vary overtime (Nguyen et al. 2021). The model by Kumbhakar (1990) and Battese and Coelli (1992) have several common features. Both models impose distributional assumptions on the statistical noise and inefficiency component and employ the maximum likelihood estimation. Besides, both models permit the mean of firm effects to vary systematically with time. One basic difference between Kumbhakar (1990) and the Battese and Coelli (1992) models is that the former has one more parameter to be estimated. The basic model runs as:

$$Y_{it} = f(x_{it};\ \beta) exp\ (v_{it} - u_{it})\ \ i = 1, 2, \ldots \text{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

$u_{it=}G(t)$. u_i under Kumbhakar (1990) or $\eta_{it}u_i$ under Battese and Coelli (1992). u_i is the time invariant component of inefficiency randomly distributed as i.i.d. non-negative truncation of $N(\mu, \sigma_u^2)$ distribution. The time varying part of the model, $G(t)$, equals $\{1 + \exp(\gamma t + \delta t^2)\}^{-1}$ under the Kumbhakar (1990) model and $\eta_{it}u_i = \{exp\ [-\eta(t - T])\}u_i$ under the Battese and Coelli (1992) model. T refers to the terminal period of the panel. $\gamma, \delta\ and\ \eta$ are an unknown scalar parameter to be estimated.

The Kumbhakar (1990) model restricts the temporal pattern of the inefficiency to be the same for all the firms in the sample (Belotti, 2013). In the model γ and δ are additional parameters to be estimated that would be used to test time invariance of technical inefficiency by setting both equal to zero ($\gamma = \delta = 0$). Similarly, in the Battese and Coelli (1992) model η represents the hypothesis about the evolution or steadiness of individual inefficiencies over the period of the study. The exponential specification of the behavior of the firm effects overtime is a rigid parameterization in that technical efficiency must either increase ($\eta > 0$), decrease ($\eta < 0$) or remain constant ($\eta = 0$). A positive value of the parameter implies increase in efficiency over time. Time invariance of inefficiencies in the Battese and Coelli (1992) model can be tested by setting ($\eta = 0$). When the scalar parameter equals one ($\eta = 0$), the Battese and Coelli (1992) collapses to the random effects maximum likelihood model of Schmidt and Sickles (1984). The Battese and Coelli (1992) model incorporates unbalanced panel data associated with observation on a sample of N firms over T time periods. This is possible by making the T^{th} period error (last time period‟s error), U_{iT} to equal U_i, that is at t=T $\{exp\ [-\eta(t - T])\}u_i$ collapses to U_i. Thus, the parameters μ and σ^2 are estimated for the last period of the panel.

In the Kumbhakar (1990) and Battese and Coelli (1992) models the intercept is the same across all firms. This generates a misspecification bias in the presence of time invariant unobservable factors, which are unrelated with the production process but that affect the output. The technical inefficiency is indistinguishable from unobserved individual heterogeneity that is the technical inefficiency confounds with the time-invariant unobserved individual effect (Nguyen et al, 2021). Therefore, to the extent the effect of these unobserved factors is captured by the inefficiency term, the model produces biased results (Belotti, 2013). Furthermore, these models do not distinguish between time invariant (persistent) and time varying (transient) inefficiencies (Kumbhakar et al, 2015). Splitting the magnitude of persistent inefficiency from the transient inefficiency is important because it reflects the effects of inputs like management as well as other unobserved inputs which vary across firms but not over time (Kumbhakar et al, 2015).

Consequently, recent efforts have been made to develop models that disentangle persistent inefficiency from firm effects. Kumbhakar et al (2015) identified two major improvements of these models from the time-varying models. First, unlike the previous models these models take factors having persistent effect on efficiency into account. Second, the time-varying models assume a firm‟s inefficiency at time t is independent of its inefficiency in the previous period. This seems unlikely as firms learn overtime to eliminate their inefficiencies by removing some of its short time rigidities; some sources of inefficiencies will stay with the firm overtime.

The most notable models that separate persistent and transient inefficiencies are Colombi et al (2014) and Kumbhakar et al (2014). Both these models develop a function that divides the error terms into four components. The first component captures firm effects; the second component captures the time-varying inefficiency; the third captures persistent inefficiency; and the fourth component is the random shock or noise. These models distinguish latent heterogeneity with long run inefficiency (Colombi et al, 2014). The model is specified as:

$$y_{it} = a_0 + f(x_{it};\beta) + \mu_{it} - u_{it} - \eta_i + v_{it} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

$$\mu_{it} \sim iid\ N(0, \sigma_\mu^2)$$

$$u_{it} \sim iid\ N^+(0, \sigma_u^2)$$

$$\eta_i \sim iid\ N^+(0, \sigma_\eta^2)$$

$$v_{it} \sim iid\ N(0, \sigma_v^2)$$

In the model $u_{it} > 0$ and $\eta_i > 0$, truncated or half normal distributions, are time-varying and persistent inefficiencies respectively while μ_{it} and v_{it} are firm effects and noise. u_{it} and η_i are truncated or half normal distributions. This model is particularly appropriate when the panel period is moderately long (Colombi et al, 2014). Although, the model specifications by Colombi et al (2014) and Kumbhakar et al (2014) are the same, the estimation approaches are different. Colombi et al (2014) estimates the above model using a single stage maximum likelihood method whereas Kumbhakar et al (2014) uses a three-step procedure. The maximum likelihood estimate by Colombi et al (2014) provides a more efficient and less biased measurement (Nguyen et al 2021) while the multi-step procedure by Kumbhakar et al (2014) is simpler and easier to implement. Recent researches have applied a four error term model. Ivan et al. (2019) and Badunenko and Kumbhakar (2017) have employed a model that separates the transient (time-varying) and persistent inefficiency among the European banks and Indian banks respectively.

All the above approaches to measuring efficiency have been employed in bank efficiency researches. The Kumbhakar et al (2014) and Colombi et al (2014) models that measures both persistent and transient inefficiencies provide better basis to understand the dynamics of banks‟ efficiency over time. Understdning the distiniction between transient and persistent inefficiencies is fundsmental to understanding bank performance. These models separate and thence assist to analyse inefficiencies due to structural factors and fundamental capabilities and due to short term variations. By distinguishing between the two components of inefficiency, it helps policy makers and researchers to better understand the drivers of inefficiency. In addition, these models resolve the misspecification bias from time invariant unobservable factors in the Kumbhakar (1990) and Battesse and Coelli (1992) models. Finally, comparatively the Kumbhakar et al (2014) model is more easily applicable than the Colombi et al (2014) model. Therefore, in this study I have applied the model suggested by Kumbhakar et al (2014).

3.4. Specification of the Cost Function

Parametric frontier analysis methods require specifying the functional form for the cost, profit or production relationships among inputs, outputs and environmental factors (Berger and Humphrey, 1997). In the efficiency analysis of banks the transcendental logarithmic (translog) function and Fourier flexible specifications have widespread applications (Girardone et al, 2004). The bank efficiency studies using parametric method have adopted either the translog or the Fourier flexible cost functions.

Both the translog functional specification and the Fourier flexible functional forms are commonly used in bank efficiency measurements. Studies that compare the efficiency measurements under the two methods provide comparable results. For instance, Berger and Mester (1997) found essentially the same results using a translog and Fourier flexible specifications, a difference of 1% on the cost efficiency.

3.5. Control variables

The arguments that need to be included in the efficiency function are crucial for the accurate measurements of a bank‟s efficiency. Two important variables standout in this regard: quality of assets and level of capital. Mester (1996) have stated that the bank efficiency studies ignore the impact of risks and quality of assets on banks‟ costs and suggest that asset quality measures and risk characteristics need to be incorporated in the underlying cost function because unless asset quality and risk are controlled for the banks‟ level of efficiency might be miscalculated.

3.5.1. Bank Inefficiency and Quality of Assets

High nonperforming loans or poor asset quality could imply the use of lower resources than usual in the initial credit evaluation and monitoring of its loans. Thus, it may seem to have higher level of efficiency, at least in the short term, when compared to banks spending resources to ensure the quality of their loans. This higher efficiency may stay until such time that loan defaults later materialize and increase the bank‟s operating costs through collection efforts and participating in the liquidation of collaterals (Resti, 1997). Efficiency measurements have to consider nonperforming loans into account. Otherwise, the unaccounted difference in the asset quality of banks could be incorrectly measured as difference in cost efficiency (Berger and Mester, 1997). Several researchers such as Mester (1996); Mitchel and Onvural (1996); Berger and Mester (1997) Atlunbas et al (2000) and Giradone et al (2004) emphasized controlling for quality of assets in efficiency measurement on grounds that their exclusion results in miscalculation of X-inefficiencies and economies of scale.

Non-performing loans could result from bad macroeconomic conditions and economic shocks, the bad luck hypothesis; from deliberate strategy to reduce credit evaluation and monitoring costs, the skimping hypothesis or from poor practice in underwriting, monitoring, and control of loans, the bad management (Berger and DeYoung, 1997). Using a granger-causality analysis, Berger and DeYoung (1997) attempted to determine the causal relationship between cost efficiency of banks and nonperforming loans under the three hypotheses. The result indicates that bad management dominates skimping at an industry level while skimping is exhibited at individual banks. In addition, they found out that bad luck persists.

Conceptually, inclusion of nonperforming loans as arguments in the cost function of banks depend on its exogeneity to banks. Berger and DeYoung (1977) argued that non-performing loans caused by bad management and skimping are endogenous to banks and should not affect the cost frontier. Thus, only nonperforming loans due to bad luck are exogenous and need to be included in the frontier functions to control for asset quality. Nonetheless, in practice, separating nonperforming loans due to bad management is problematic. In effect, most research include nonperforming loans as an argument in the efficiency functions (Hughes and Mester, 1993; Mester, 1996; Hughes et al, 1997; and Girardone et al, 2004). There are variations in the measurement of quality assets under this category. Berg, Forsund, and Jansen (1992) (cited in Berger and Mester, 1997) has used loan losses instead of nonperforming loans in measuring quality of assets.

3.5.2. Financial Capital (Equity) and bank efficiency

The financial capital affects a bank‟s costs directly or indirectly. Indirectly, the amount of financial capital determines a bank‟s insolvency risk which depends on a bank‟s capacity to absorb portfolio losses. In turn, higher solvency risk increases a bank‟s cost through higher risk premium on uninsured deposits, intensity of risk management activities, and discount rate applied to future benefits (Berger and Mester, 1997; Hughes and Mester, 2008). There is a tradeoff between the extent of solvency risk assumed and the amount of a bank‟s financial capital. The level of a bank‟s financial capital could indicate the risk-preference of the bank. Risk averse banks will have larger level of capital than that minimizes costs (Berger and Mester, 1997). Therefore, efficiency of risk averse banks will be mismeasured even though they have behaved optimally given their risk preference (Berger and Mester, 1997).

Directly, financial capital can affect bank costs by providing alternative to deposits as source of funding for loans (Mester 1996, Berger and Mester 1997; Hughes and Mester, 2008). In economic terms the cost of equity is higher than the cost of debt due to its high required rate of return. Efficiency measures commonly use accounting costs that exclude all costs except capital raising costs related to financial capital while taking all costs related for debt (deposits). In effect, the impact of financial capital on efficiency depends on the relative size of capital raising costs and the cost of debt. When capital raising costs dominate the costs of deposits, banks that do not use financial capital for loan funding will have lower costs and better efficiency score than banks using financial capital; and when the costs of deposit dominate capital raising costs, banks using financial capital as a source of credit funds will have lower costs and better efficiency (Berger and Mester, 1997).

Therefore, the underlying industry cost function for bank efficiency measurement should incorporate capital to control for the risk characteristics of banks and changes in the cost structure between banks that use financial capital to fund their loans and those that do not. Inclusion of financial capital in efficiency measurement controls for the risk preference of banks and the variation due to relative costs of debt and equity (Mester, 1996; Berger and Mester, 1997). Consequently, several studies have included financial capital in measuring efficiency of banks (Hughes and Mester, 1993; Clark, 1996; Mester 1996; Berger and Mester, 1997; Atlunbas et al. 2001; Girardone et al. 2004; Mosunda, 2008). Clark (1996) found out that including equity in the cost efficiency function raised measured cost efficiency. Berger and Mester (1997) found similar results for profit efficiency. This improvement in cost efficiencies could be due to better fit of the data and the ensuing reduction in the specification error that otherwise could be counted as inefficiency (Maudos, 1996a).

3.5.3. Technical Progress and bank efficiency

Technical progress refers to the reduction in the cost function of banks over time. Technical progress allows the firm to produce a given output, *Q*, at lower levels of total cost over time, holding input prices and regulatory effects constant (Atlunbas et al, 2001). Technical progress shifts the cost frontier of banks (Bikker and Bos, 2005). The reduction in the cost function of the banks could be brought about through learning by doing, organizational changes and changes in regulations (Atlunbas et al., 2001). The rate of technical progress may be inferred from changes in the firm's cost function over time. When panel data is used to evaluate efficiency, the changes in the frontier due to technological progress has to be controlled for.

Efforts to measure the effect of technology on banks have initially been studied by Hunter and Timme (1991) who assessed the effect of technology on large US banks. Subsequently, Lang and Welzel, 1996; McKillop et al.,1996; Atlunbas et al., 1999; Atlunbas et al., 2001; Molyneux, 2003; and Mosunda, 2008 have attempted to measure the effect of changes in cost efficiency due to technological progress. Technical progress is accounted as a function of time-trend that is assumed to represent all the changes allowing for more efficient use of existing inputs, together with the effect of other factors like changing environmental conditions of banks through time (Baltagi and Griffn, 1988 cited in Atlunbas et al., 2001). Thus, technical progress is commonly measured as an elasticity of total costs with respect to time i.e. $\partial TC/\partial t$ (Lang and Welzel, 1996; KcKillop et al., 1996). However, instead of such single blanket measure of technical progress, the Ivan et al. (2019), following Baltagi and Griffin (1988) and Heshmeti and Kumbhakar (1996) (both cited in Ivan et al, 2017), has adopted a method that classifies technical progress in to three "pure-technological progress", "scale-augmenting technological progress" and "non-neutral technological progress". This method takes technical progress to result from exogenous non-economic time trend and other exogenous economic factors, called technology shifters. Regardless, the measurement that takes time trend as a catch all measure of technical progress is common in bank efficiency measurements.

Empirical results provide evidence of cost reductions from technical progress on banks. Technical progress accounted for a 2.8% - 3.6% per year cost reduction in European banks between 1989 and 1996 (Atlunbas et al., 1999). A study by the Ivan et al. (2019) obtained a fairly close result on European banks for 2006 to 2017. On the other hand, Hunter and Timme (1991) found an average annual technical progress of 1% among large US banks for the period between 1980 and 1986. KcKillop et al., (1996) found a reduction in the real cost of large Japanese banks for the period 1978 to 1991. Generally, larger banks earn larger cost reduction from technical progress than small banks (Hunter and Timme, 1991; Atlunbas et al., 2001, Atlunbas et al.1999). Exceptionally, Lang and Welzel (1996) found higher cost reduction for smaller banks than larger banks for Cooperative banks in Germany. They found technical progress rate of 1.4% for the largest banks and 2.4% for the smallest banks. Technical progress could also be affected by bank ownership. Badunenko and Kumbhakar (2017) have found different patterns of technical progress for state owned, private, and foreign banks in India.

In summary, theroretically quality of bank assets, proxied by bad loans, is both exogeneous and endogeneous to the bank. Only bad loans due to bad luck are exogeneous to a bank and need to be included to control for quality of assets. Intentional decision by a management, skimping, and imprudence, bad management, are endogeneous to a bank and have to be subsumed in the bank efficiency measure. However, separating exogeneous bad loans from that of endogenous bad loans is impractical. In effect, as controlling for quality of assets provides more accurate results, incorporating asset quality in the measurement model is advisable. Financial capital have to be also included as a control variable in measuring efficiency of banks because it reflects the bank‟s risk management, strength and cost structure and regulatory compliance. All these affect a bank‟s costs directly and indirectly. Failure to control for financial capital would subsume the cost structure effects of financial capital in the inefficiency measurement. Inclusion of financial capital would enable more accurate measure of a banks efficiency. Finally, including the time variable is of paramount importance to understand the improvements or declines in the operational efficiency of banks through time.

3.6. Exogenous Inefficiency Determinants

In the stochastic frontier studies, the main interest is not only to estimate and predict technical inefficiency but also to identify the source of inefficiency and estimate the marginal effects of these determinants (Lai and Kumbhakar, 2018). Policy measures can be taken when the factors giving rise to inefficiency among the firms is identified. Thus, identifying sources of inefficiency is as crucial as measuring the level of inefficiency. In effect, stochastic frontier models, apart from estimating inefficiencies attempt including exogenous variables have been proposed. The exogenous variables, neither inputs nor outputs of the production process but nonetheless affect the productive unit performance, have been incorporated into the stochastic frontier models in a variety of ways. Efforts to identify the covariates of efficiency have been proposed by Pitt and Lee (1981); Battese and Coelli (1995); Ivan et al. (2019) and Badunenko and Kumbhakar (2017). Except for the difference in the efficiency measure types, these models attempted to identify determinants. Pitt and Lee (1981) modeled the time invariant inefficiency against the time invariant factors; Battese and Coelli (1995) regressed the exogenous variables against the time varying inefficiency using a single-step maximum likelihood approach. Recently, Badunenko and Kumbhakar (2017) and Ivan et al. (2017) employed a pooled OLS approach to identify the sources of both persistent and transitory/transient inefficiency.

3.7. Model Specification of the Study

This study fundamentally employs the stochastic cost frontier method following Aigner et al (1977) and Meeusen and Van De Broeck (1977). I have employed panel data approach because efficiency is better studied and modeled with panels (Cornwel et al, 1990; Kumbhakar, 1993). In this model the cost function for i^{th} firm is represented by the model $TC = TC\,(Q_i\,, P_{\mathrm{j}}, B) + \varepsilon_i$; where TC is the observed total cost of production for bank i, Q_i is the vector of bank outputs for bank i, P_{j} is the vector of input prices for bank j and B is a vector of parameters. In this model the error term ε_i is a two-component error term that contains statistical noise and inefficiency. It can be represented as $\varepsilon_i = v_i + u_i$; where v_i is a two-sided error term that is assumed to be independently and identically distributed and u_i is a non-negative random variable representing inefficiency.

Specifically, I have employed the four error component model proposed Kumbhakar et al (2014). Kumbhakar et al (2014) allows separation of the composite error term from the stochastic noise; individual effect, persistent inefficiency and time-varying or transient inefficiency. The Kumbhakar et al (2014) adopted a four components error model. Kumbhakar et al (2014) used a three-step procedure. The basic feature of this model is that it allows decomposing persistent bank effects into random bank-specific effect, as proposed in Green (2005), and persistent technical inefficiency. Thus, this model decomposes the error of the stochastic cost function into four components: the time-varying (transient) inefficiency; time-invariant (persistent) inefficiency; a bank-specific effect that captures heterogeneity across banks; and pure random component as in Green (2005).

In this study I have applied the fixed effects model proposed by Kumbhakar et al. (2014). Accordingly, the stochastic frontier model employed in this study can be written as:

$$\ln TC_{it} = a_0 + lnTC(y_{it}, w_{it}, \theta; \beta) + \mu_i + v_{it} + \eta_i + u_{it} \qquad \text{.......................................}(8)$$

Where a_0 is a constant, TC_{it} is observed total cost of bank i at time t, y_{it} are outputs of a bank, w_{it} are the input prices, θ are bank specific control variables, β are a vector of parameters, μ_i is bank specific effect, v_{it} is random disturbance or noise, u_{it} is a time - varying inefficiency measure and η_i is persistent (time-invariant) inefficiency. Ratio of non-performing loans and total capital to total assets ratio are the bank specific control variables used in the model.

The stochastic frontier for bank costs is the sum of the constant, random disturbance and the bank specific effect. The difference between the observed cost and the stochastic frontier represents inefficiency. Following Kumbhakar et al. (2014) the above equation can be written as:

$$\ln TC_{it} = a_{0*} + lnTC(y_{it}, w_{it}, \theta; \beta) + a_i + \varepsilon_{it} \quad (9)$$

Where $a_{0*} = a_0$ - E(η_i) - E(u_{it}), $a_i = \mu_i - \eta_i +$ E(η_i) and $\varepsilon_{it} = v_{it} + u_{it}$ - E(u_{it}). With this specification a_i and ε_{it} have a mean of zero and constant variance. As specified by Kumbhakar et al. (2014) this model has three steps. As this is a familiar panel data model, in the first step the standard fixed effects panel regression is used to estimate β and to predict a_i and ε_{it}. This step provides predicted values of a_i and ε_{it}; $\hat{a}_i$ and $\hat{\varepsilon}_{it}$. The second step estimates the transient or time varying efficiency, u_{it}, from $\hat{\varepsilon}_{it}$. The predicted value, $\hat{\varepsilon}_{it}$, from the first step on $\varepsilon_{it} = v_{it} - u_{it} +$ E(u_{it}) is employed. This step estimates the time varying inefficiency based on Jondrow et al. (1982). The third step estimates the persistent inefficiency, η_i, using a similar process. For this I used the best linear predictor of a_i from step one using $a_i = \mu_i - \eta_i +$ E(η_i) by assuming μ_i is $i.i.d.\, N(0, \sigma^2_\mu)$ and η_i is $i.i.d.\, N^+(0, \sigma^2_\eta)$ and estimated using the standard normal-half normal stochastic frontier model cross-sectionally and obtain estimates of the persistent technical inefficiency components, η_i, using the Jondrow et al. (1982) procedure. The overall technical efficiency is the product of persistent and time-varying efficiencies.

With regard to specification, I have used a translog specification. In the model I have included time trend, financial capital, and non-performing loans to capture technological process, risk and relative cost of debt and equity, and quality of loans. In this framework pursuant to the intermediation approach banks are modeled as to produce loans and other earning assets using labor, physical capital, and deposits. With this specification the above general equation can be presented as follows:

$lnTC_{it} =$

$$a_0 + \sum_{i=1}^{2} \alpha_i lnQ_i + \sum_{j=0}^{3} B_j P_j + \tau_k lnS, + \tau_s lnK + \phi_t T + 1/2[\sum_{i=1}^{2} \sum_{i=1}^{2} \delta_{ij} lnQ_i lnQ_j + \sum_{i=1}^{3} \sum_{j=1}^{3} \gamma_{ij} lnP_i P_j + \tau_{kk} lnSlnS + \tau_{ss} lnKlnK + \phi_{tt} T^2] + \sum_{i=1}^{2} \sum_{j=1}^{3} \rho_{ij} lnQ_i lnP_j + \sum_{i=1}^{2} \alpha_{is} lnQ_i S + \sum_{i}^{2} \alpha_{it} lnQ_i T + \sum_{j}^{3} \alpha_{jt} lnP_j T \sum_{j=1}^{3} \beta_{js} lnP_i S + \sum_{i=1}^{2} \alpha_{ik} Q_i K + \sum_{j=1}^{3} \beta_{jk} P_j K + \tau_{ks} lnSlnK + \varepsilon_i \quad (10)$$

Where TC is a measure of the total cost of production comprising operating costs and financial costs; Q_1 and Q_2 are output quantities, that are total loans and advances and other earning assets, respectively. Other earning assets are financial assets of banks other than loans and advances that mainly include investments, treasury bills and deposits made in other banks. P_1 is the price of labor computed as total personnel cost divided by the average number of employees; P_2 is the price of fixed capital computed as operating costs other than personnel costs divided by the total fixed assets; P_3 is the price of deposits computed as total interest expense divided by total deposit liabilities; K is the level of financial capital; S is non-performing loans; $\alpha, \beta, \rho, \gamma, \lambda, \delta, \theta, \phi$ *and* τ are parameters to be estimated and ε_i is a four component error term.

Financial capital is measured as the capital asset ratio Girardone et al. (2004) or total value of capital (Mester, 1996, Berger and Mester, 1997, Atlunbas et al. 2001) or as a quasi-fixed input (Hughes and Mester, 2008). In this study I have used total value of financial capital. Capital appreciations from valuations for IFRS adoption by banks are excluded from assets and financial capital. Most studies use nonperforming loans to measure quality of assets. Berg, Forsund and Jansen (1992) (cited in Berger and Mester, 1997) has used loan losses instead of nonperforming loans in measuring quality of assets. In this study I have used the ratio of nonperforming loans to total loans to measure the quality of assets.

A time trend variable, T, is included to track rate of technical change throughout the panel period. Similar to Lang and Welzel (1996) a time trend variable is used to serves as a proxy for technical change. The impact of technical change is the variation in the average cost due to a change in technology measured by the partial derivative of the estimated cost function with respect to the time trend (Land and Welzel, 1996; Atlunbas et al, 2001; Mosunda, 2008). That is:

i j

$$\frac{\partial lnTC}{\partial T} = \phi_t + \phi_{tt} T + \sum^2 \alpha_{it} lnQ_i + \sum^3 \alpha_{jt} lnP_j \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

In addition, standard symmetry has to be imposed on the translog portion of the function: $\delta_{ij} = \delta_{ji}$ and $\gamma_{ij} = \gamma_{ji}$ where $(i = 1,2)$ *and* (j = 1,2,3) and the following linear restrictions are necessary for linear homogeneity in factor prices $\sum_{j=1}^{3} \beta_j = 1; \sum_{i=1}^{3} \gamma_j = 0; \sum_{j=1}^{3} \rho_{ij} = 0; \sum_{j=1}^{3} \beta_{jk} = 0; \sum_{j=1}^{3} \beta_{js} = 0$. In accordance with the assumed constraint of linear homogeneity in prices, TC, P_1 and P_2 are normalized by the price of capital, P_3.

For the determinants of inefficiency, bank specific characteristics have been regressed against the inefficiency measures with a pooled regression using the following model:

$$u_{it} = z_{it}\,\delta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

$$\eta_i = z_{it}\,\delta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13)$$

Where u_{it} is the time varying efficiency is measure; η_i is the time invariant efficiency measure; z_{it} are bank level characteristics and δ are parameters. The variables employed the year the bank is established that measures the age of the bank. Age could be related to efficiency as bank operation could involve learning by doing (Berger and Mester, 1997). Return on asset (ROA) measured by the ratio of net income over total assets to capture impact of profitability; branch networks measured by the total number of branches the bank has at the end of the year taken from National Bank reports; total asset of the bank measures with total assets of the bank; intermediation ratio the extent to which the bank converts deposits to loans and is measured as the ratio of total deposits to total loans. Finally, I have included the ratio of non-core deposits to total financial assets ratio (or non-physical assets of the bank). Non-core deposit implies purchased funds or deposits other than demand and saving deposits. This variable measures reliance of banks on purchased funds that is the difference among banks in terms of the extent to which they rely on their non-core deposit to extend their loans. Reliance on non-core deposit relates to efficiency since the cost of purchased funds differs from core deposits.

3.8. Data sources and Sample

The data for this study contains an unbalanced panel data of seventeen Ethiopian private banks from 1998 to 2021. The banks included are those that have been operational by the beginning of 2021. The data for the study is taken from annual consolidated reports of the banks, NBE reports and the WDI data base.Data from the financial statements are extracted based on the IMF manual Annual (IMF, 2016). Averages are computed by adding the beginning amount and ending amount, and then divided by two. For banks in their first year of operation only ending values are taken as an average. All financial values are measured in terms of constant 2010 prices. For this purpose, the consumers‟ price index from the World Development Indicators (WDI) is employed. Besides, all values (except those expressed in percent) are log transformed.

Chapter Four Results and Discussion

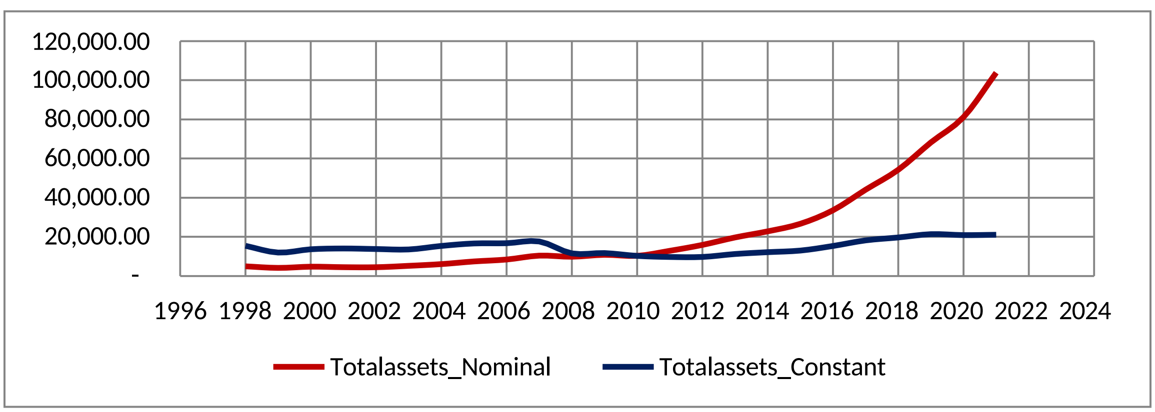
4.1. Descriptive Analysis

4.1.1. Bank Size: Total Assets, Fixed Assets and Capital

4.1.1.1. Total Assets

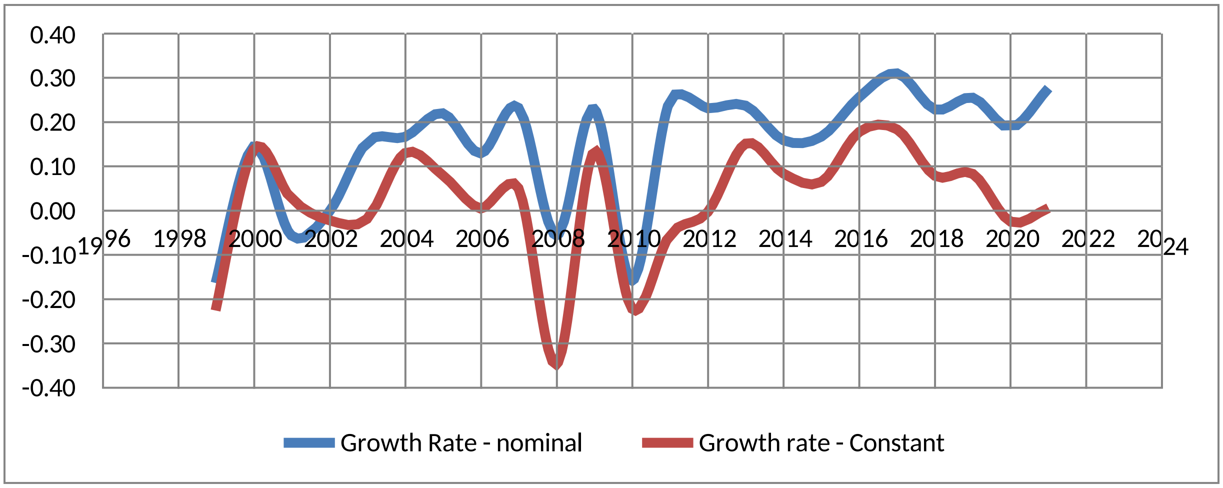
Total assets of a bank usually measure the size of a bank. The aggregate value of total assets of the seventeen banks considered in this study has shown a steady growth over the study period. The aggregated total assets of the seventeen banks has grown about 19.8 fold, from 4.97 billion in 1998 to more than 103.82 billion in 2021. In particular, the aggregate value of total assets has exhibited a sharp increase after 2010. The total asset of banks in constant prices has increased by just 36 percent from 15.45 billion to 21.08 billion from 1998 to 2021. The growth rate of total assets in current prices has a high annual growth over the years and has been positive except for four years. The highest growth rate of assets of banks in nominal values is 31 percent in 2017, whereas the lowest growth rate is negative sixteen percent in 1999. On the other hand, except for 2008 and 2017, total assets in constant values of banks is generally characterized by smaller changes than total assets in nominal values. The highest and lowest growth rates of total assets in constant value during the study period are 19 percent and negative thirty four percent in 2017 and 2008 respectively. There is a sharp decline in the aggregate value of bank assets from 2008 to 2011. This could mainly be due to a sharp surge in the national CPI in those years; 44. 39 percent and 32.01 percent in 2008 and 2011 respectively.

Figure 1: Total Assets of banks from 1998 -2021 (current and constant values)



Source: *Own computation from Annual Report of banks*

Figure 2: Growth of total assets (Current and constant values)

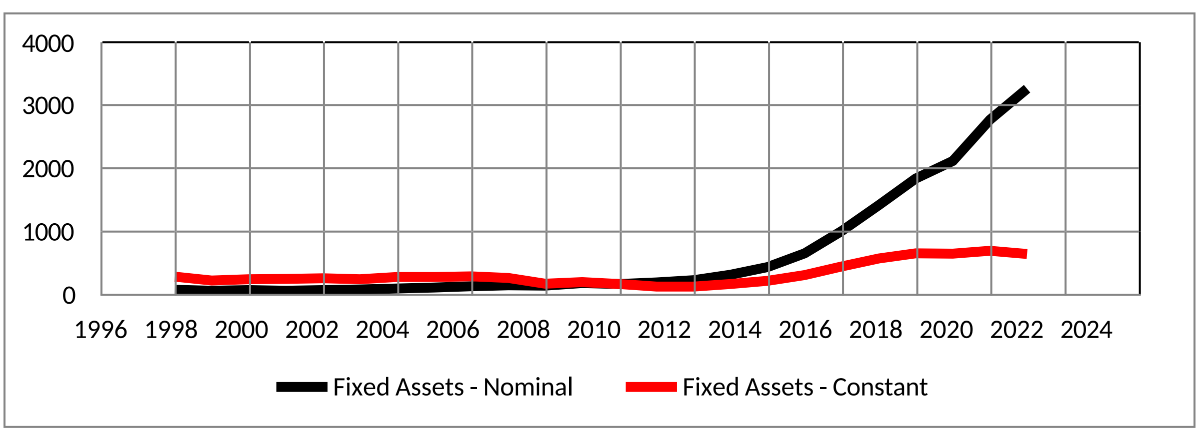


Source: *Own computation from Annual Report of banks*

4.1.1.2. Fixed Assets

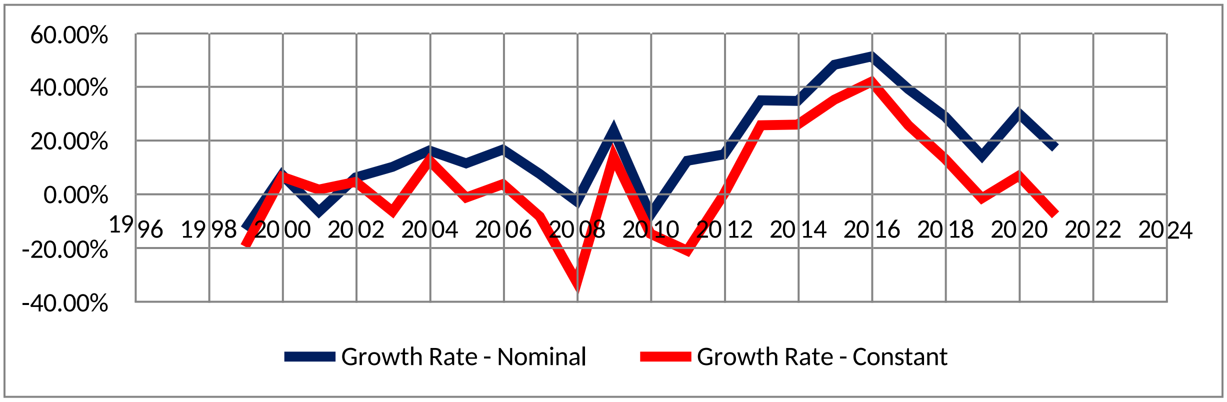
The pattern of fixed assets of the banks for the period 1998 to 2021 is different from the pattern of total assets. Unlike total assets, the fixed assets of banks have remained the same until 2013. From 2013 onwards the average total fixed assets in current prices of the banks has exhibited a sharp increase. Similarly, the value of fixed assets in constant prices has shown a high growth rate after 2013. The growth rate of fixed assets seems to decline after 2016. The fixed assets in current prices of all the banks have increased by 32.4 fold from 98.17 million in 1998 to 3.2 billion in 2021. The value of fixed assets in constant prices, on the other hand, has increased by 1.18 fold from 304.9 million to 665.8 million.

Figure 3: Fixed asset of banks 1998-2021 (current and constant prices)



Source: *Own computation from Annual Report of banks*

Figure 4: Growth of fixed assets of banks 1998-2021 (current and constant prices)

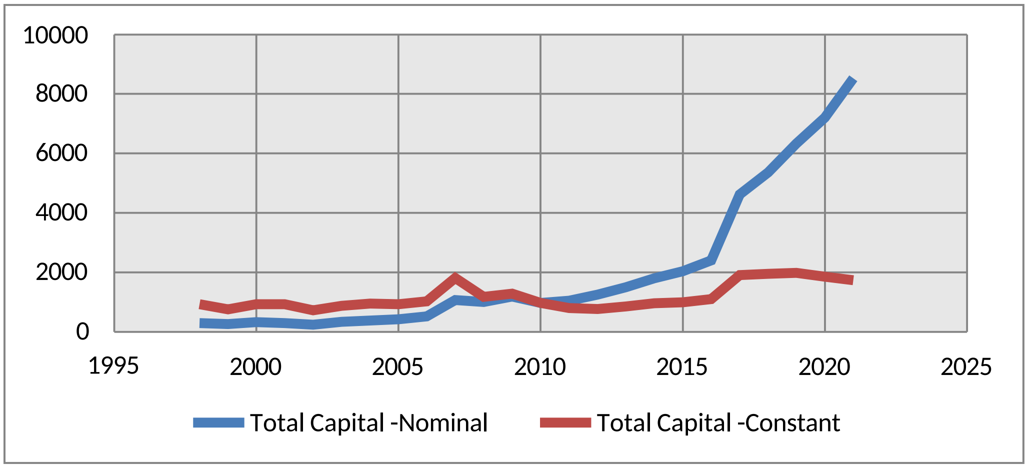


Source: *Own computation from Annual Report of banks*

4.1.1.3. Capital

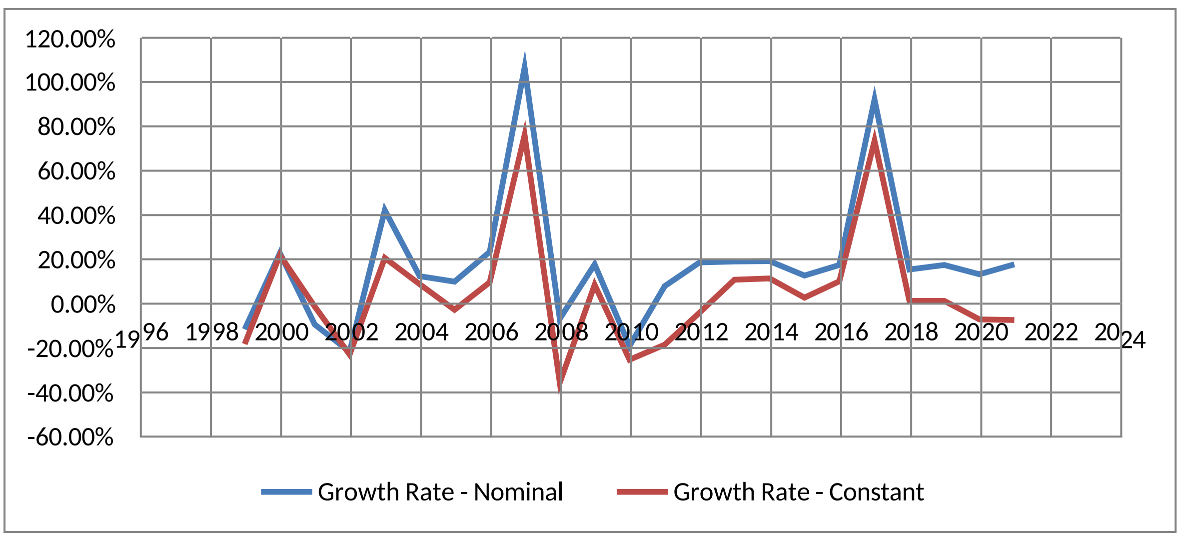
Banks are required to increase their capital amount to ensure capital adequacy. The change in the capital of the banks has initially been very slow. The nominal capital of banks has shown a slow growth up to 2010. Then the growth rate has picked up and gotten high from 2010 to 2015. After 2015 the capital of banks has, further, shown an even higher and steady growth. The highest and lowest growth rates of the aggregate capital of banks in current prices are 106.92% in 2007 and negative 21.05 percent in 2002 respectively. The nominal value of total capital of banks increased by 27.6 fold from 297.3 million to 8.5 billion. The capital value in constant prices has a more stable growth pattern. In total, it has increased by 87% over the period 1998 to 2021. The highest growth rates of the aggregate capital is 76.55 percent registered in 2007. On the other hand, the lowest growth rate of capital of aggregate capital is negative 35.76 percent in 2008; probably due to the sharp increase in CPI in that year. By 2021 the aggregate capital of the banks in current and constant prices has reached over 8.5 billion and 1.7 billion respectively. The

Figure 5: Total capital of banks from 1998-2021 (Current and constant prices)



Source: *Own computation from Annual Report of banks*

Figure 6: Growth rate of capital (current and constant prices)



Source: *Own computation from Annual Report of banks*

The inter-bank comparison of total assets, fixed assets and capital of banks reveals that Commercial Bank of Ethiopia is still by far the largest bank in terms of total assets, fixed assets, and capital. The bank with the smallest average total assets is Debub Global Bank (Now renamed Global Bank Ethiopia). The variation in total asset, fixed asset and capital among banks is very high. Another striking fact is that the share of the Commercial Bank of Ethiopia, though declined slightly, has remained very high even with the establishment of so many banks.

Table 1: Average fixed assets, average total assets & average capital of banks (in millions)

Bank	Fixed Assets		Total Assets		Total Capital	
	Nominal	Constant	Nominal	Constant	Nominal	Constant
AB	988.29	421.81	22,782.67	10,095.25	2,844.57	1,193.33
AbB	260.80	87.00	9,448.88	3,008.70	1,433.94	465.90
AdIB	118.22	41.20	3,528.95	1,156.75	753.49	257.35
BIB	284.28	83.76	9,245.32	3,029.42	1,438.89	484.42
BOA	913.01	313.92	14,847.06	6,641.83	1,581.96	744.45
BuB	272.31	85.23	8,436.37	2,792.68	1,333.82	452.66
CBE	4,565.83	2,054.24	226,698.00	110,824.50	14,109.15	6,777.68
Coop	706.51	225.02	24,134.12	7,603.07	2,225.05	723.29
DB	1,022.28	432.40	20,056.74	10,074.04	2,297.20	1,057.80
GBE	127.43	37.00	4,670.07	1,325.87	825.02	241.99
EB	144.11	45.86	6,652.99	2,052.66	1,008.90	313.46
LB	248.50	76.90	9,812.11	3,356.42	1,242.78	458.41
NIB	715.02	244.09	12,710.31	5,877.27	1,819.86	868.36
OIB	561.94	195.37	14,789.58	4,908.47	1,881.42	632.35
UB	600.82	209.93	11,890.05	5,221.84	1,407.66	644.11
WB	867.45	325.35	17,081.59	7,097.10	2,587.73	1,126.69
ZB	364.35	106.74	8,519.18	2,796.28	1,348.31	431.04
Overall						
Mean	970.41	391.36	33,060.14	15,197.91	2,872.59	1,265.50
Std. Dev.	2,462.92	756.36	106,226.50	35,341.34	6,931.45	2,286.29
Min	3.08	6.78	88.23	253.51	21.70	37.02
Max	20,115.71	5,270.02	991,319.00	223,471.10	53,826.54	18,418.35

Source: *Own Computation from Annual Reports of Banks*

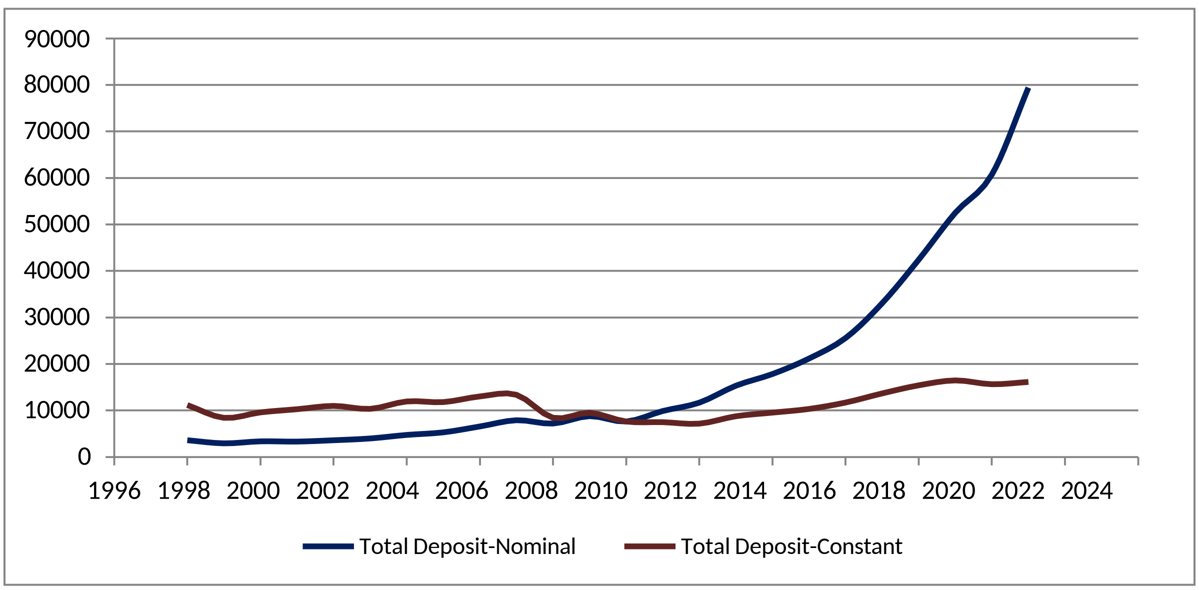
4.1.2. Resource Mobilization

Mobilization of resources or deposits from customers is a key function of banks. Under the views of the intermediation approach, deposits are the fundamental inputs for bank outputs; loans, and earning investments. The resource mobilization of banks is also one of the major sources of expenses for banks. Banks incur significant costs in order to mobilize deposits and to serve their depositors. Banks incur substantial costs of labor, fixed, and finance costs to mobilize deposits. Thus, efficiency in resource mobilization may translate to higher over all efficiency. In effect, success in resource mobilization is one of the determinants of bank efficiency. The resource mobilization of banks has to be analyzed from two perspectives that are gross mobilization of deposits and the ratio between the different types of deposits that is core, and non-core deposits.

The core deposits of banks are demand and saving deposits while the non-core deposit is the fixed time deposit. These two types of deposit imply different mobilization costs.

As indicated in Figure 7, nominal value of deposits mobilized by banks has shown substantial growth over the study period. In particular, the growth in total deposits has become higher and sharper after 2011. On the other hand, when adjusted for CPI, the deposit mobilized by banks, more or less, stays constant. In fact, CPI adjusted deposits have had a declining trend from 2007 to 2012. The decline during this period could be the increase in the national inflation rate during the period which have depleted the deposits of banks. This change is ,some how, in tandem with the decline in bank assets and capital during the same period. From 2012 onwards, the CPI adjusted deposits has again recorded some growth. The current value of total deposits increased by 21.08 fold from Birr 2.9 billion to Birr 79.4 billion whereas the constant value increased by 91.2% from Birr 8.4 billion to Birr 16.12 billion.

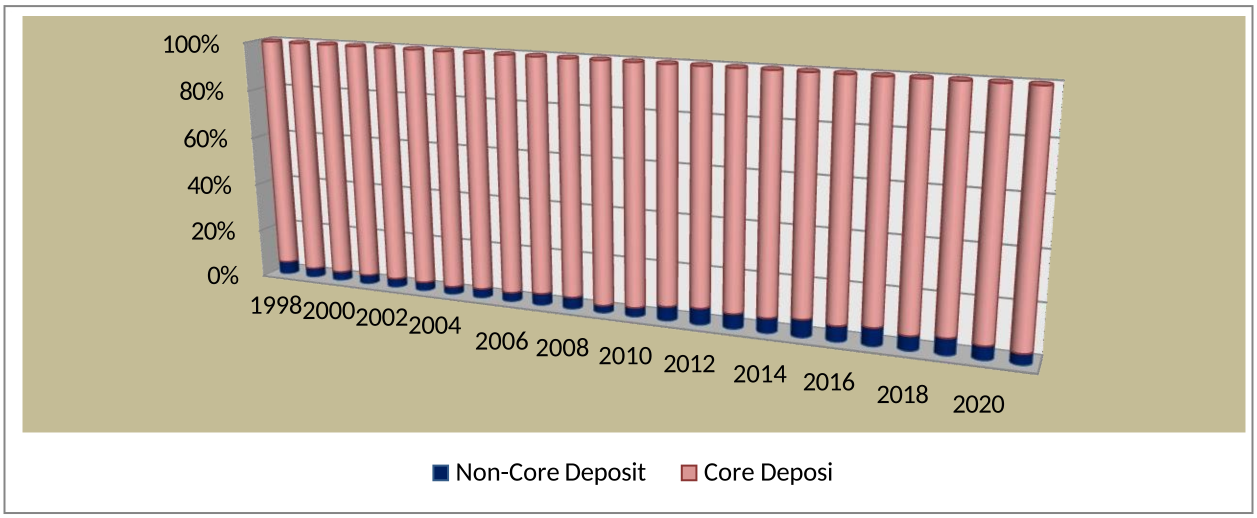
Figure 7: Total value of average deposits 1998 -2021



Source: *Own computation from Annual Report of banks*

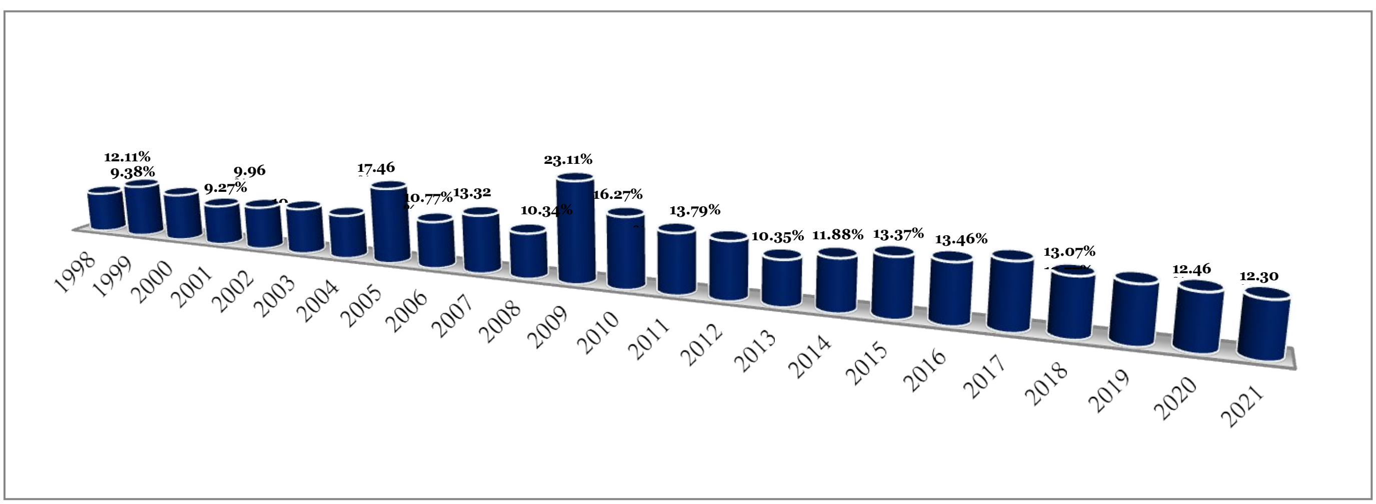
An important issue in resource mobilization among banks is the share of non-core deposits from the total deposits of banks. The core and non-core mix of deposits affects the cost and cost structure of banks because the core and non-core deposits cater for different cost structures. While core deposits incur more labor and fixed capital costs to serve depositors and find customers, non-core deposits require comparatively larger finance costs. Normally, interest rates for non-core deposits are set through negotiations. The ratio of non-core deposits to total deposits of banks has not changed significantly from 1998 to 2021. The pattern, however, has been irregular; from 1998 to 2009 the ratio of non-core deposits has an overall increasing trend after reaching its maximum ratio in 2009 (23.11%), then the core deposit ratio has a decreasing trend up to 2013 after which the ratio increases up to 2017. Since, 2017 the non-core deposit ratio remains stable.

Figure 8: Composition of core and non-core deposit of banks 1998-2021



Source: *Own computation from annual reports of banks*

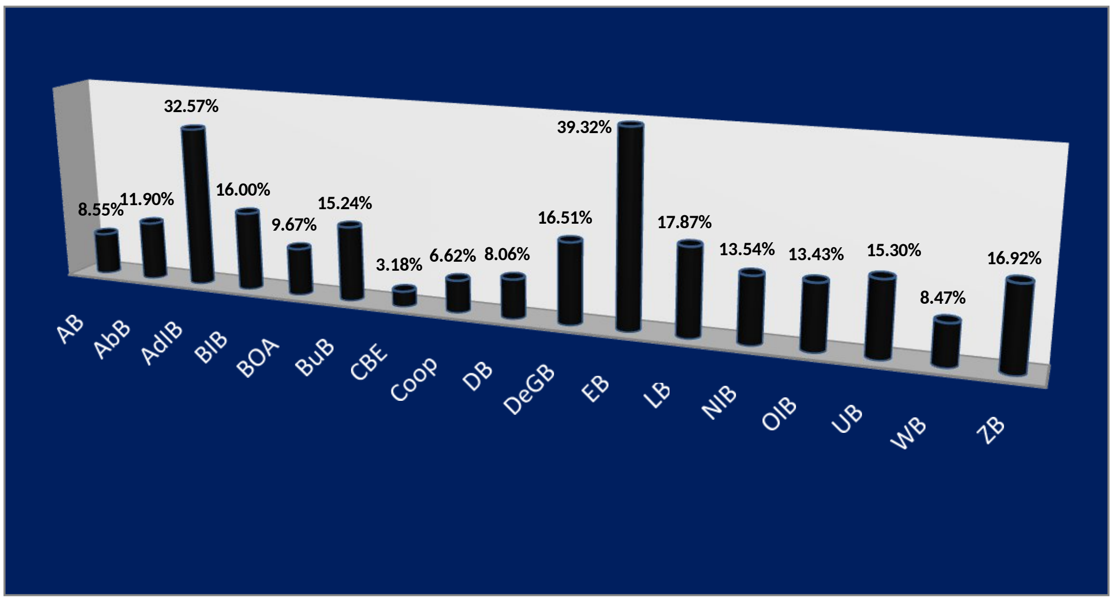
Figure 9: Ratio of Non-core deposits to total deposits of banks from 1998 - 2021



Source: *Own computation from annual reports of banks*

Comparison of non-core deposit ratio among banks reveals that Enat bank (39.32%), Addis International Bank (35.57%), and Lion Bank (17.87%) are the first three banks with the highest non-core deposit ratios. Zemen Bank (16.92%) and Global Bank Ethiopia (16.51%) also have comparatively high non-core deposit ratios. Commercial Bank of Ethiopia (3.18%), Cooperative Bank of Oromia (6.62%), and Dashen Bank (8.06%) have the lowest non-core deposit ratios. Wogeagen Bank (8.47%) and Awash Bank (8.55%) also have low non-core deposit ratios. Younger and smaller banks have more non-core deposits ratio indicating their higher dependence on non-core deposits than older and bigger banks.

Figure 10: Non-core deposit ratio by banks



Source: *Own computation from annual reports of banks*

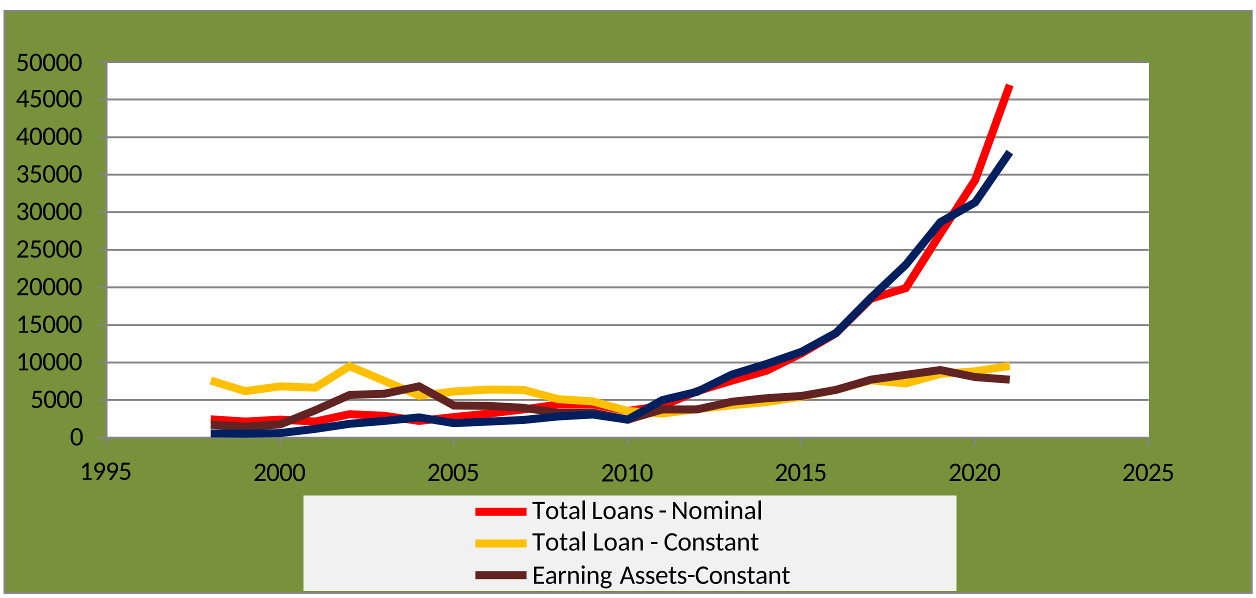
4.1.3. Bank Outputs and Quality of Loan

This study takes loans and advances of banks as outputs of banks. The amount of loans is substantially larger than investments in earning assets such as shares, Treasury Bills, and bonds. The portion of earning assets has increased after 2011 mainly due to the Ethiopian Government Bond issued for construction of the Renaissance Dam. Banks were required to use 27% new loans they extend to buy the Government Bond. The interest rate on this bond is 3.25% and does not earn much income for the banks.

The total loans and advances of banks have shown very little growth from 1998 to 2010. However, from 2010 onwards loans and advances have a very fast growth. Probably, this could be due to the entry of several new banks after 2010. Nonetheless, the CPI adjusted value of loans has not changed over the study period. The total loans and advances in current prices have increased by 18.2 fold from 1998 to 2021 that is from Birr 2.45 billion to Birr 46.97 billion. While the loans and advances at constant prices have increased by 25.56 percent from 7.60 billion to 9.54 billion over the same period.

Earning assets are sources of income for banks other than loans and advances. These assets mainly refer to investment on Treasury bills, government bonds, and shares of other companies. As indicated in Figure 11 below, up to 2010 the total earning assets, other than loans, grow at a very low rate. However, after 2010 the growth rate increased and the total of earning assets increased steadily. The total earning assets of the banks seems to move in tandem with the value of loans and advances. The similarity in the pattern of movement between earning assets and loans could imply that banks do not take these two outputs as substitutes. In fact, both earnings and loans increase parallel to deposits which implies their pattern is determined more by the amount of available deposits than a decision made based on the comparative earnings of loans and earning assets.

Figure 11: Loans, Earning assets and Non performing loans of banks 1998-2021

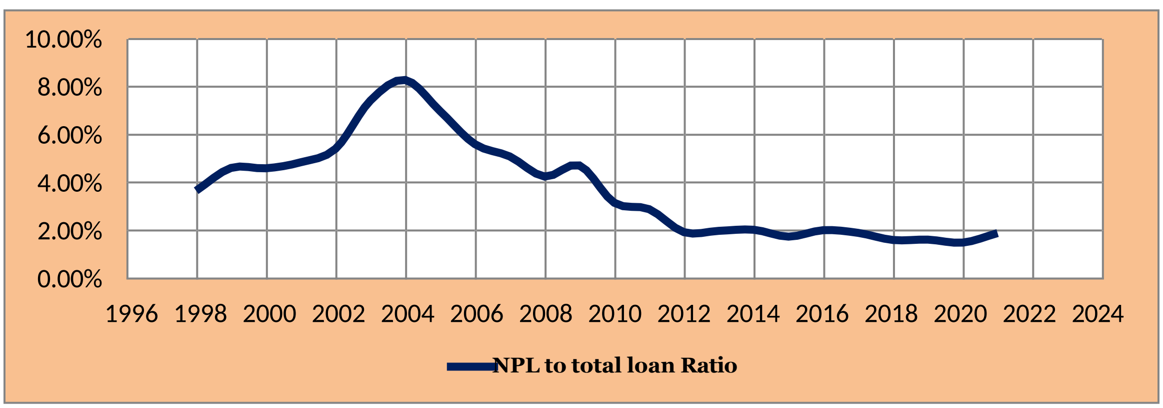


***Source**: Own computation from annual reports of banks*

Apart from their amount, the quality of the loans and advances is of paramount importance to banks. The ratio of non-performing loans to gross loans is a measure of loan quality. From 1998 to 2021, the ratio of non-performing loans to gross loans, or the NPL ratio, has a dichotomous pattern. From 1998 to 2004, the NPL ratio increased sharply and reached its peak of 8.33% in 2004. After 2004, the NPL ratio declines steadily and by 2021 reached 1.94% of total loans. The general trend of the quality of bank loans as indicated by the NPL ratio generally has an increasing trend from 2004 to 2021.

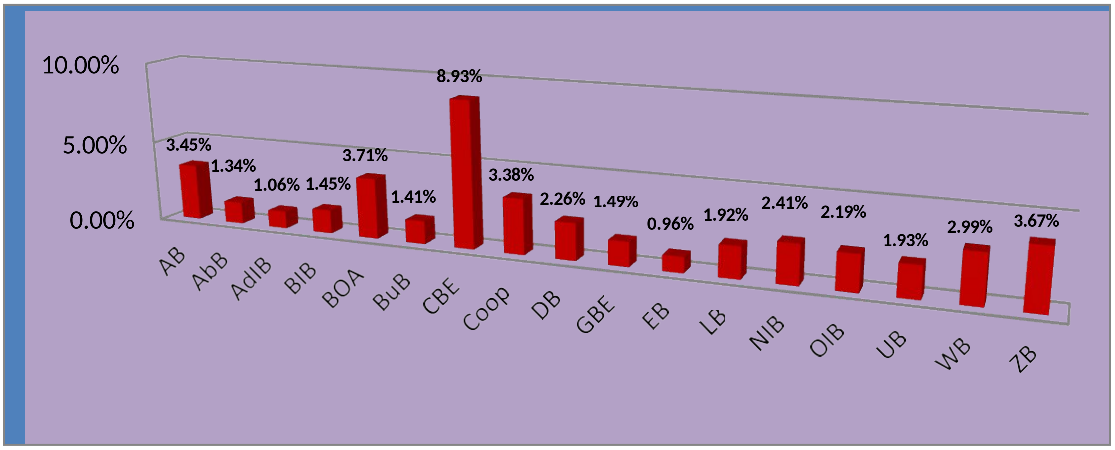
Comparing banks in terms of loans and loan quality reveals that Enat Bank (0.96%), Addis International Bank (1.06%), Abay Bank (1.34%), and Bunna International Bank (1.41%), respectively, have the lowest average NPL ratio; whereas Commercial Bank of Ethiopia (8.93%), Bank of Abyssinia (3.71%) and Zemen Bank (3.67%) and Awash Bank (3.45%) have the highest NPL ratio. Except for Zemen Bank, all the banks with the highest NPL ratio are first generation old banks; conversely, banks with the lowest NPL ratio are younger and smaller banks.

Figure 12: The ratio of non-performing loans to total loans and advances 1998-2021



Source*: Own computation from annual reports of banks*

Figure 13: Non Performing Loan ratio of banks



Source*: Own Computation from annual reports of banks*

4.1.4. Costs and Cost Structure

4.1.4.1 Bank Costs

In addition to size, resource and outputs, an important aspect of banks‟ operational performance relates to costs and their structure or mix. Structurally, the cost of banks can be classified as operating costs and finance costs. The operating cost of banks is comprised of salary and benefits, and fixed capital costs such as rent and depreciation. Table 2 below presents the total costs, operating costs, and finance costs of banks from 1998 to 2021. The total cost of banks in constant prices has increased by 194% or from Birr 475.6 million to arround 1.4 billion for the period 1998 to 2021; whereas at current prices the total costs has increased by 44.4 fold from 153.14 to 6,788.37. During the same period, operating costs have grown by 312% from Birr 171.5 million to Birr 704.73 million; operating costs at current prices have increased by 62.1 fold. Finally, finance cost of banks have a smaller increment than operating costs; it increased by 127% and 32.9 fold in constant and current prices for the period 1998 to 2021 respectively. The variation in all these costs as measured by standard deviation and range of the costs is very high.

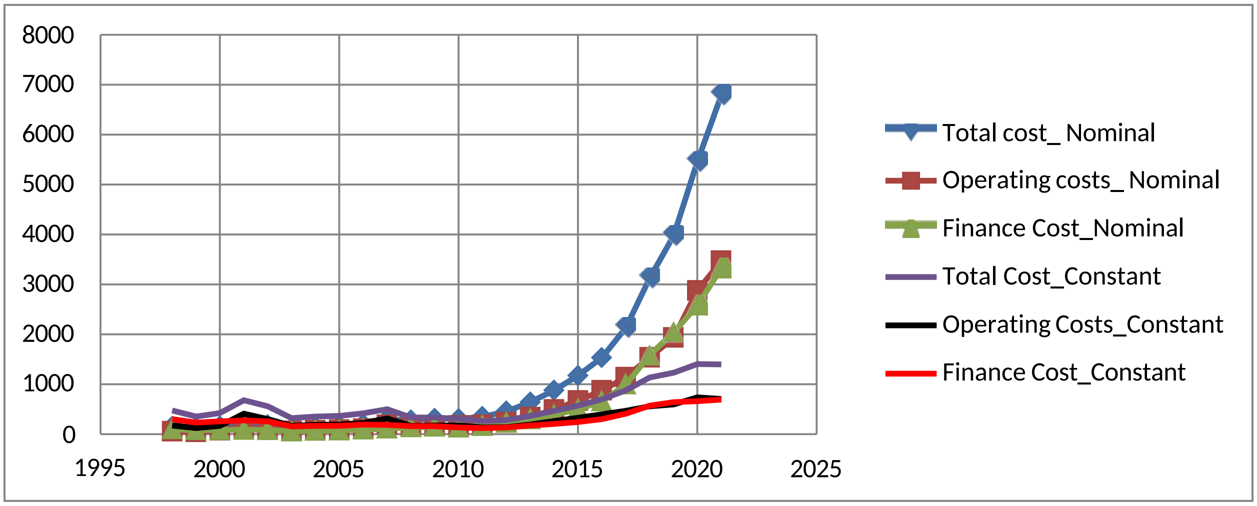
Table 2: Total, operating and finance costs of banks from 1998-2021 (in millions)

Year	Total Cost		Operating Costs		Finance Cost	
	Constant	Nominal	Constant	Nominal	Constant	Nominal
1998	475.59	153.14	171.47	55.21	304.12	97.93
1999	354.77	123.37	125.23	43.55	229.54	79.82
2000	426.66	160.62	170.77	76.01	255.89	84.61
2001	682.38	219.63	408.56	131.51	273.82	88.12
2002	560.54	183.29	302.43	98.89	258.11	84.4
2003	319.54	122.91	167.03	64.25	152.51	58.66
2004	356.76	141.64	190.09	75.47	166.67	66.17
2005	367.71	164.82	196.47	88.06	171.24	76.76
2006	415.58	200.1	224.49	107.84	191.09	92.26
2007	504.21	297.72	314.99	186.00	189.22	111.72
2008	331.48	282.69	170.94	145.78	160.54	136.91
2009	336.82	311.53	172.7	159.73	164.12	151.8
2010	306.6	306.60	172.78	172.78	133.82	133.82
2011	263.76	348.17	137.93	182.07	125.83	166.1
2012	281.64	458.97	143.97	233.13	137.67	225.84
2013	366.51	641.40	194.89	341.06	171.62	300.34
2014	469.51	878.46	263.97	493.89	205.54	384.57
2015	571.47	1,171.52	328.36	673.13	243.11	498.39
2016	700.52	1,529.72	400.58	873.91	299.94	655.81
2017	882.3	2,134.43	471.11	1,139.71	411.19	994.72
2018	1,133.23	3,120.15	558.43	1,537.45	574.8	1,582.7
2019	1,234.51	3,972.49	595.08	1,932.69	639.43	2,039.8
2020	1,404.48	5,451.68	740.23	2,873.28	664.25	2,578.4
2021	1,396.13	6,788.37	704.73	3,470.08	691.4	3,318.29
Overall						
Mean	681.15	1,761.57	355.29	915.06	326.12	846.51
Std. Dev.	1,578.54	5,660.52	752.42	2,639.91	836.06	3,047.17
Min	9.36	3.26	7.42	2.58	0.84	0.68
Max	12,710.68	58,333.07	6,353.11	25,190.08	6,730.91	33,142.99

***Source**: Own Computation from annual reports of banks*

Figure 13 presents the trend of total costs, operating costs, and finance costs over the period 1998 to 2021 under both current prices and constant prices. As indicated in the figure, all the costs have a similar pattern of growth. In terms of current prices all the costs have small growth until 2011 and a higher growth rate after wards. Especially, the nominal value of costs increased sharply after 2014. The costs under constant prices changed irregularly, however, similar to the costs in current prices , they grow steadily after 2015.

Figure 14: Trend of total, operating and finance costs (current and constant prices)

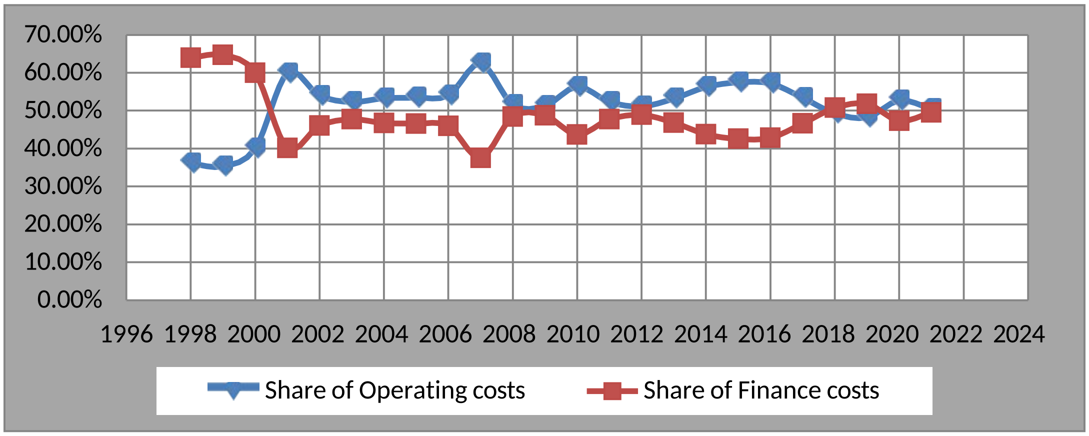


Source: *Own Computation from annual reports of banks*

4.1.4.2 Structure of Costs

The structure or mix of operating and finance costs of banks provides information on the relative importance of these costs in the operation of banks. Figure 15 presents the composition of total costs. Initially, until 2001 the finance costs had a larger proportion of the total costs. As depicted in figure 15, the gap in the proportion of finance costs and operating costs is the highest in 1998 and 1999. The gap, however, has reduced sharply and in 2001 the proportion of operating costs surpassed that of the finance costs. From 2001, all the way to 2017 operating costs have higher proportion than the finance costs. After 2017 the proportion of the two costs has become balanced and they more or less have equal proportions all the way to 2021.

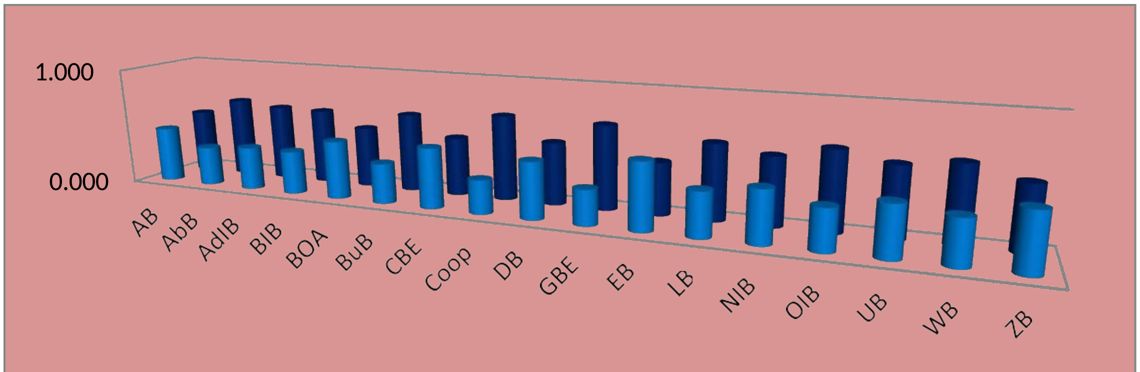
Figure 15: Trend of the share of finance costs and operating costs 1998 - 2021



Source: *Own computation from Annual Report of banks*

The inter-bank data shows that a difference in the share of finance and operating costs to total costs among banks. The result shows that for all the banks, except Enat Bank and Commercial Bank of Ethiopia, the average share of operating costs is higher than that of the finance costs. The case for Commercial Bank of Ethiopia (CBE) may need further assessment. Probably, the reason for lower operating costs ratio for CBE could be related to its ownership. As it s the government‟s policy bank much of its deposit is mobilized without much effort on resource mobilization. Among the banks, Cooperative bank of Oromia, Global bank of Ethiopia and Abay bank have the highest operating cost proportions and the highest disparity between the proportions of finance and operating costs. On the other hand, the Commercial Bank of Ethiopia and Bank of Abyssinia have the most balanced ratio of finance and operating costs to total costs followed by Zemen Bank, Awash Bank, and Dashen Bank.

Figure 16: Share of finance costs and operating costs of banks 1998-2021



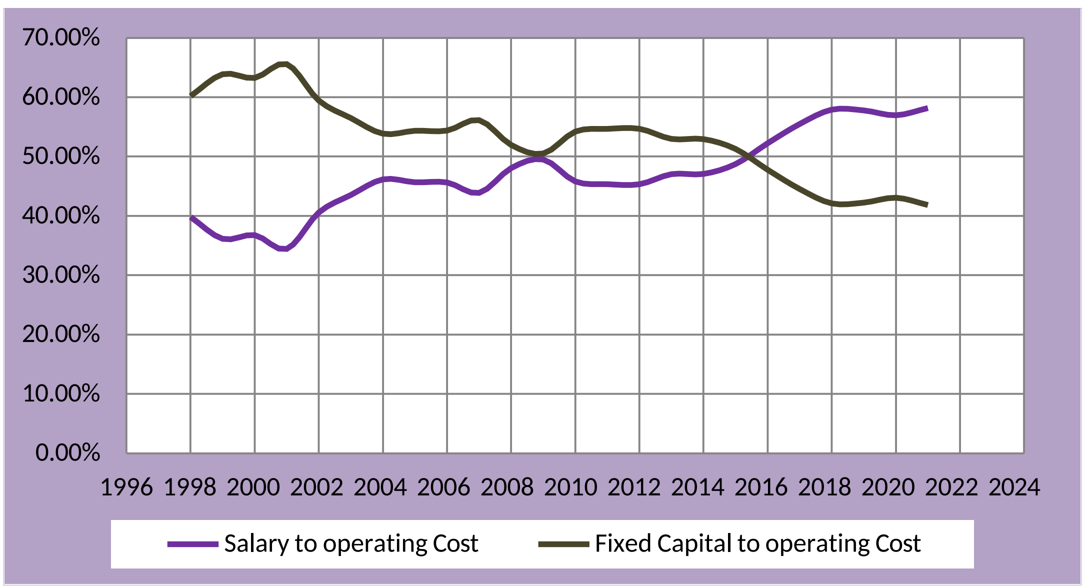
	AB	AbB	AdIB	BIB	BOA	BuB	CBE	Coop	DB	GBE	EB	LB	NIB	OIB	UB	WB	ZB
Interest Cost ratio	0.467	0.329	0.364	0.366	0.490	0.339	0.509	0.286	0.477	0.296	0.560	0.373	0.435	0.342	0.420	0.368	0.472
Operating cost ratio	0.533	0.671	0.636	0.634	0.510	0.661	0.493	0.714	0.523	0.704	0.440	0.627	0.565	0.658	0.580	0.632	0.528

Source*: Own computation from annual reports of banks*

The operating costs can be further analysed in terms of labor costs and fixed capital cost proportions. Impact of expenditures in banking technologies on the proportion of labor and fixed capital costs is fairly clear. Investments in banking technologies increase the relative importance of fixed capital costs over labor costs through increase in depreciation and amortization costs and reduction in the size of labor. However, the impact of banking technologies could be quite insignificant until the technologies are adopted by sufficiently large number of customers. The labor cost reduction effect of technologies could be insignificant if only few customers use the technologies because increase in the adoption of technologies by customers reduces labor costs through reduction in customer visits of branches. Over the past years banking in Ethiopia is characterized by a steady build-up of fixed assets, extensive branch expansion particularly out of Addis, and adoption of banking technologies such as internet banking, mobile banking, and expansion of Automated Teller Machines (ATM). Hence, the expectation is for fixed capital costs to grow faster than labor costs. However, the data indicates a decline in the proportion of fixed capital cost proportion compared to the proportion of labor costs.

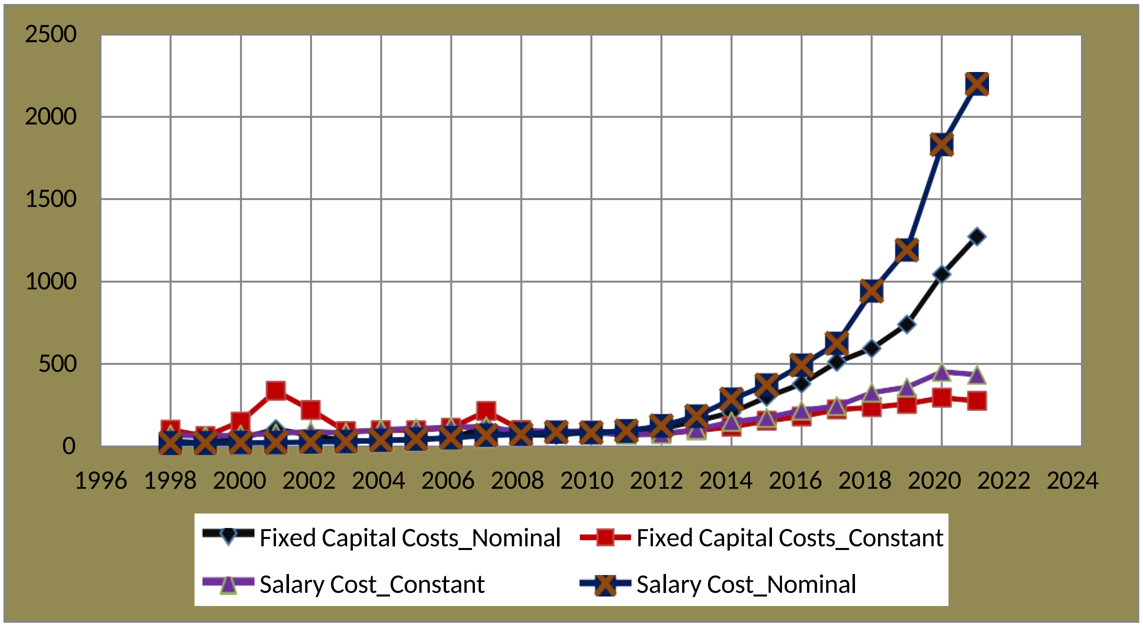
The effect of branch expansion on the proportions of labor cost and fixed capital cost is unclear. On one side, branch expansion increases the fixed capital cost of banks through rent and/or depreciation cost of branches. Concurrently, branch expansion increases the size of labor and coordination costs. Therefore, expanding branches will increase the proportion of labor costs if the additional salary and benefit costs from branch expansion is higher than the rent and other fixed costs of branches. The case of Ethiopian banks is presented in figure 17 below. In spite of aggressive branch expansion by banks and a surge in fixed assets after 2014 (presented in section 4.1.2. above) and the ensuing increase in depreciation, amortization, and rent costs; the proportion of labor costs increases while that of fixed capital costs decline through time. After 2015, the proportion of labor costs has surpassed the proportion of fixed capital costs. Based on the argument presented above, this could result from higher labor costs than rent and other fixed capital costs in branches and/or lower adoption of banking technologies by customers. Besides, these may imply an absence of technical progress among the banks regardless of investments on banking technologies.

Figure 17: Trend of salary and fixed capital cost proportions



Source: *Own computation from annual reports of banks*

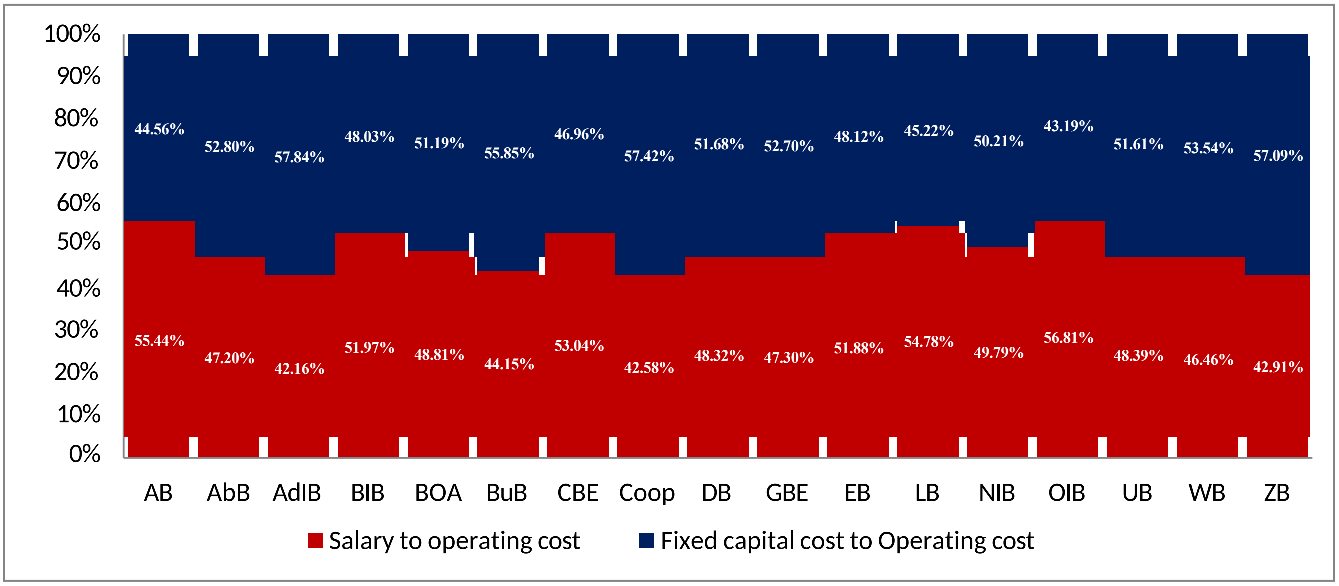
Figure 18: Trend of labor and fixed capital cost 1998-2021 (current and constant prices)



***Source**: Own computation from annual reports of banks*

The change in the proportion of labor and fixed capital costs among banks is not uniform. For some banks such as Awash bank, Birhan International bank, Commercial Bank of Ethiopia, Enat bank, Lion bank, and Oromia International bank the proportion of labor cost is higher. Among these banks, Oromia International bank (56.81%) has the highest proportion of labor costs followed by Awash bank (55.44%) and Lion bank (54.78%). Even though for most banks the proportion of labor and fixed capital costs is more or less balanced, most banks have higher proportion of fixed capital costs than labor costs. Banks with the highest proportion of fixed capital cost proportion are Abay bank (57.84%), Cooperative bank of Oromia (54.42%), Zemen bank (57.09%), and Bunna International bank (55.82%).

Figure 19: Proportion of salary and fixed capital costs to operating cost



Source: *Own computation from Annual Report of banks*

4.1.5. Price of inputs

This study follows the intermediation approach that takes labor, deposits, and fixed capital as inputs to the production of loans and earning assets. In effect, the prices of labor, deposits and fixed capital could shed some light on the operation of banks. The overall annual average price of labor is Birr 103,975.50 and Birr 53,332.00 at current and constant prices respectively. The price of labor in current prices has very high variation whereas labor prices in constant prices have moderate variation. The average price of fixed capital is Birr 0.58; the maximum is Birr 1.47, and the minimum is Birr 0.16. The average cost of finance for the entire period is Birr 0.34 per Birr of deposit or 3.4%. It has reached as high as 9.6%.

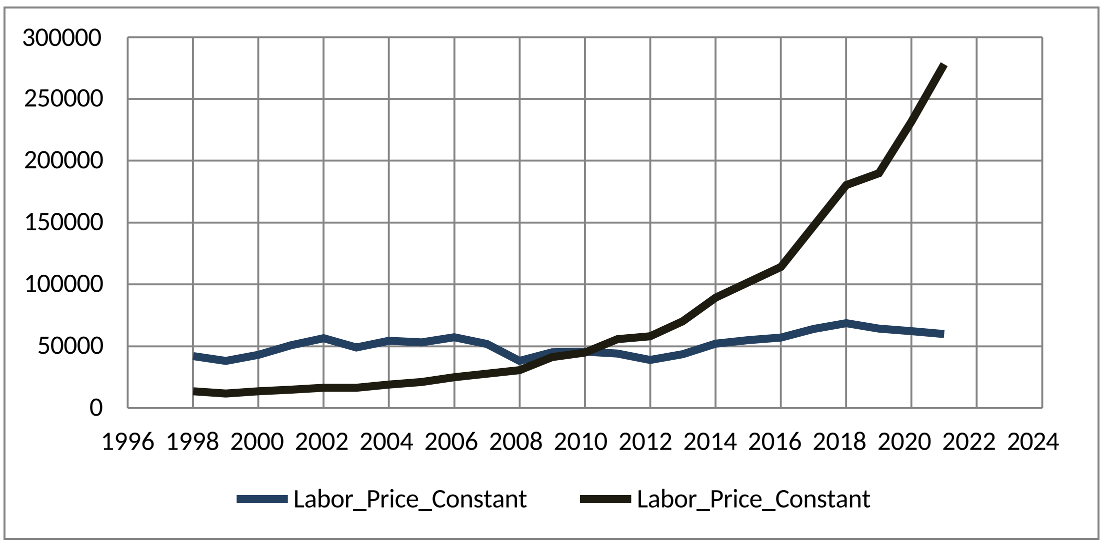
Table 3: Price of labor (current and constant prices), fixed capital and deposits

Variable	Obs	Mean	Std. Deviation	Minimum	Maximum
Labor_Price_Nominal	264	103,975.50	87,040.30	5,859.91	473,016.40
Labor_Price_Constant	264	53,332.00	15.807.97	16,838.81	114,364.30
Fixed Capital_Price	264	0.63	0.32	0.028	1.72
Deposit_Price	264	0.034	0.013	0.004	0.096

Source*: Own computation from annual reports of banks*

Figure 18 presents the trend of labor price from 1998 to 2021. The trend in the price of labor in current prices has three phases. The first phase from 1998 to 2006 is a steady but mild growth rate period. Then in 2008 the growth rate picks up and until 2016 a higher and increasing growth rate of nominal labor price is registered. Finally, from 2016 onwards the labor price of banks exhibited a very high growth rate. On the other hand, labor price in constant prices does not show significant change and remained stable for most the period. The only noticeable changes could be the decline in 2008 and the increases in 2013 and 2016.

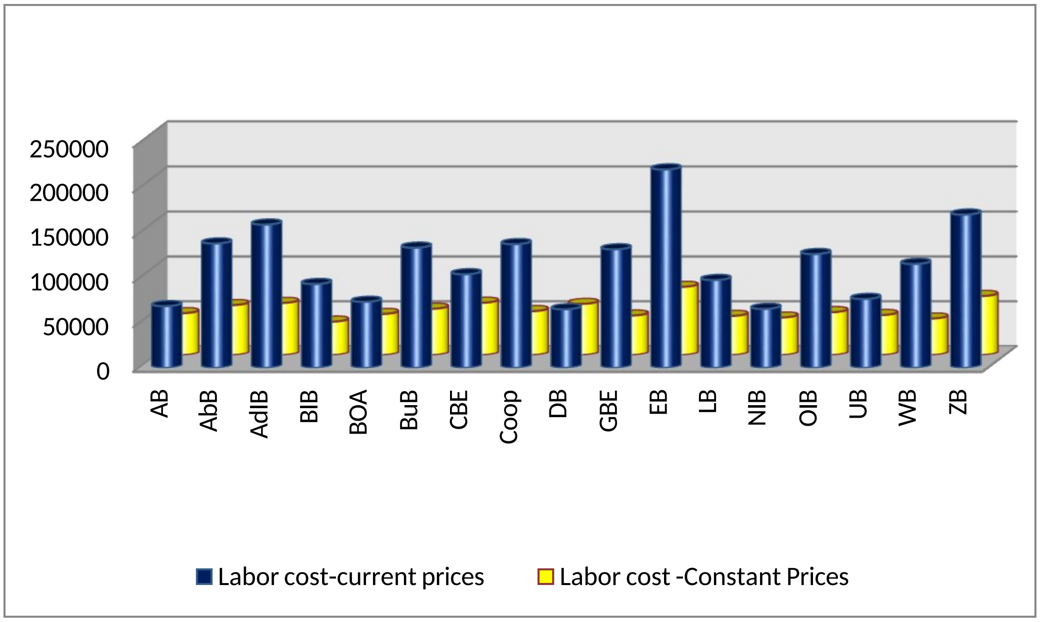
Figure 20: Trend of labor price 1998 - 2021



***Source**: Own computation from Annual Report of banks*

Comparison of the average price of labor among banks reveals that Enat bank has the highest average labor price followed by Zemen Bank and Addis International Bank. Conversely, Dashen Bank, Nib International Bank, and Awash Bank, respectively, have the lowest labor prices. Comparatively, older private banks have lower average labor price than younger private banks.

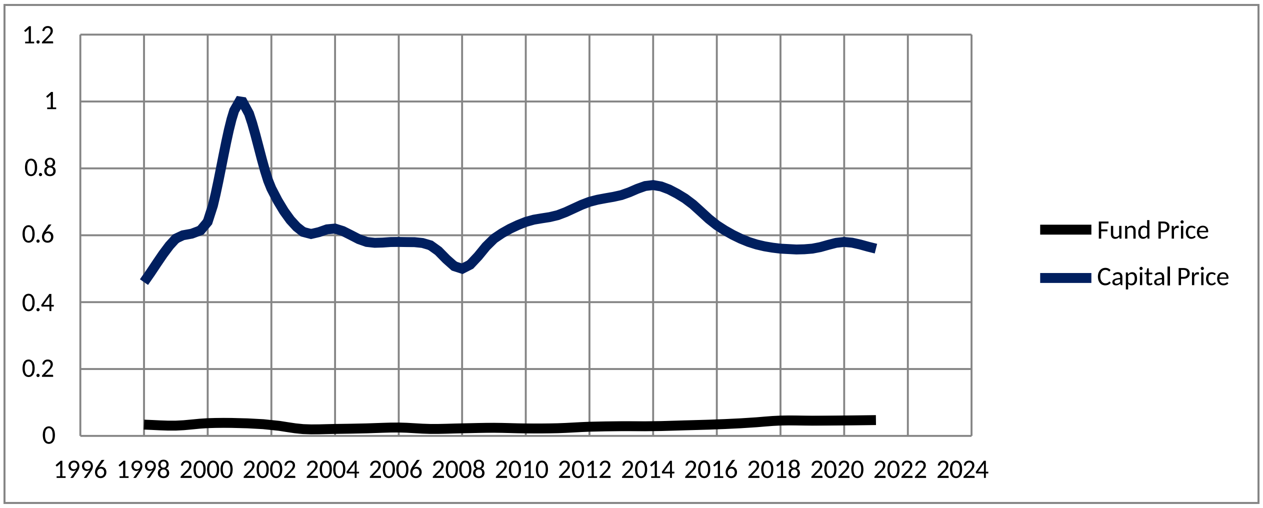
Figure 21: The average price of labor



***Source**: Own computation from Annual Reports of Banks*

Finance costs and fixed capital costs has opposing trends. Fixed capital costs have high variation across time while finance costs have low variation. The spiked graph of the fixed capital costs implies high volatility of these costs during the period. Periods of surge are followed by sharp decline. For instance, a very high increase in 2000 and 2001 is followed by a drop in 2002 and 2003. Similar but milder trends can be observed from 2007 to 2009 and 2015 to 2018. The only periods in which the fixed capital costs evinced stability are between 2004 and 2008, and from 2018 onwards. Conversely, finance costs have been stable with minor variations all along the study period. It has been close to constant over the period. The finance cost of banks is higher over the periods from 1998 to 2002 and from 2014 onwards. In fact, the finance costs have an increasing trend from 2012 onwards.

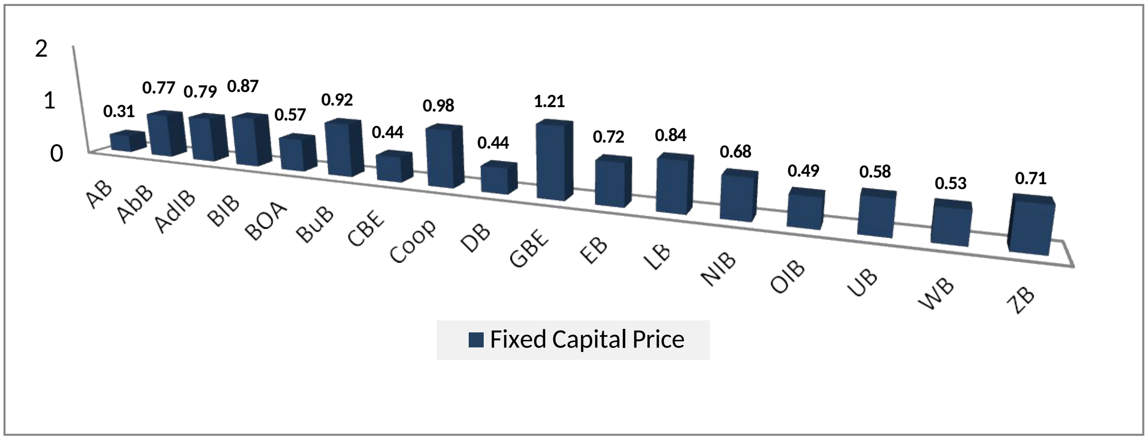
Figure 22: Trend of fixed capital cost, and finance cost of banks 1998-2021



***Source**: Own computation from Annual Reports of Banks*

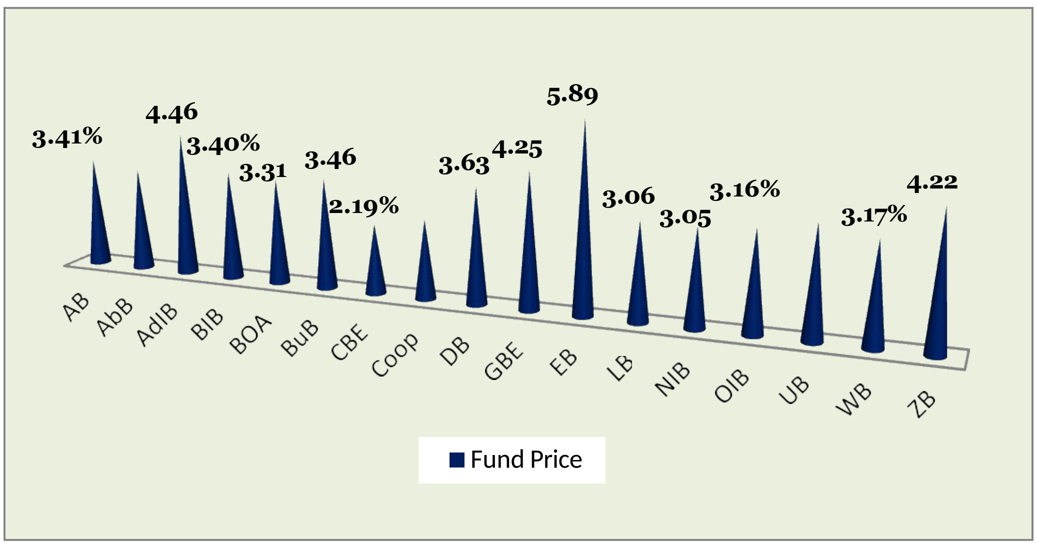
Similar to the variation across time, the variability of fixed capital costs among banks is high. The highest value is 1.46 and the lowest is 0.31. Banks with the highest average price of fixed capital are the Global bank of Ethiopia, Cooperative bank of Oromia, and Bunna International bank. Birhan Bank, Lion bank, and Abay bank also have high average fixed capital price. On the other hand, Awash bank, Commercial bank of Ethiopia, and Dashen bank have the lowest fixed capital prices. Generally, older banks have lower fixed capital price than the younger banks. Unlike, fixed capital costs, the inter-bank variation in finance costs is mild. The highest fund price is 0.0589 and the lowest is 0.0219. Banks that incur the highest finance costs per deposit are Enat bank, Addis International bank, Global bank of Ethiopia, and Zemen bank. On the other hand, banks with the lowest fund price are Commercial bank of Ethiopia, Cooperative bank of Oromia, Nib International bank, and Lion bank respectively.

Figure 23: Finance cost and fixed capital cost of banks



Source: *Own computation from Annual Reports of Banks*

Figure 24: Finance (Fund) price of banks



Source: *Own computation from annual reports of banks*

4.1.6. Profitability of Banks

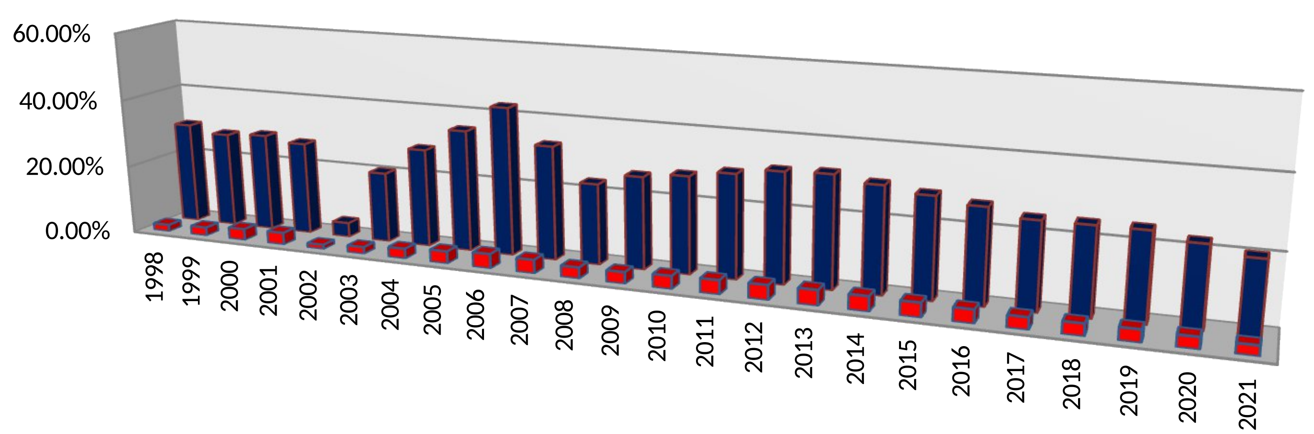
The fundamental operational performance indicator for banks is their profitability. The common measures of profitability for banks are return on equity (ROE) and return on asset (ROA). Return on equity (ROE) measures profit per Birr of owners‟ equity (capital) and ROA measures profit earned per Birr of investment on assets. To assess the profitability of Ethiopian banks ROE and ROA of banks is computed. Figure 23 presents the ROE and ROA of the banks across time. The overall average ROE and ROA of the banks from 1998 to 2021 are 26.91% and 3.55%, respectively. The highest yearly average ROE, 42.86%, was registered in 2006, while the lowest ROE is about 4.21% in 2002. The highest yearly average ROA of banks is 4.42% in 2013 and the lowest is 1.32% in 2002.

Table 4: Descriptive statistics of profitability measures

Variable	Obs.	Mean	Std. Deviation	Minimum	Maximum
ROA	264	3.55%	1.49%	-3.31%	8.31%
ROE	264	26.91%	16.16%	-47.29%	102.26%

***Source**: Own computation from Annual Reports of Banks*

Figure 25: ROA and ROE of banks 1998-2021

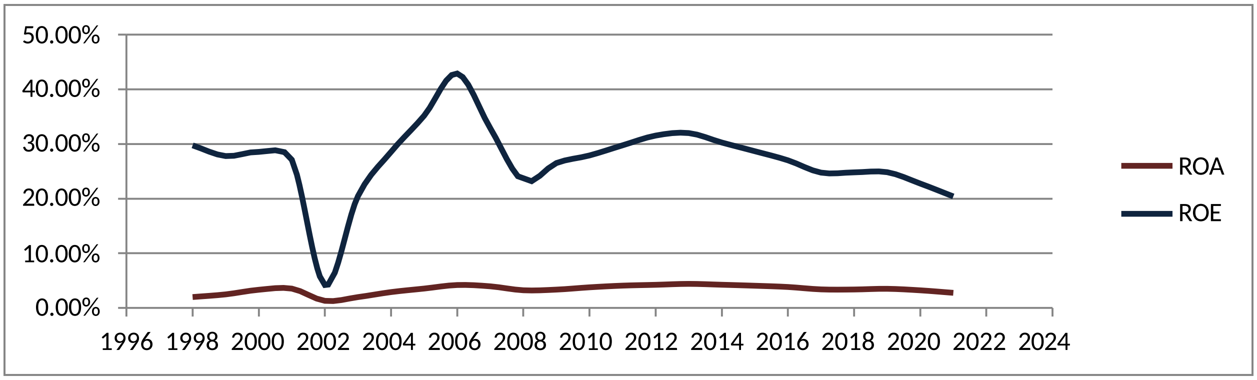


	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
ROA	2.01%	2.50%	3.34%	3.55%	1.32%	2.00%	2.92%	3.56%	4.21%	3.97%	3.25%	3.37%	3.79%	4.10%	4.25%	4.42%	4.26%	4.08%	3.85%	3.41%	3.39%	3.52%	3.24%	2.78%
ROE	29.73	27.78	28.54	27.10	4.21%	20.46	28.54	35.15	42.86	32.87	23.23	26.50	27.89	29.74	31.52	31.98	30.25	28.66	27.00	24.75	24.80	24.83	22.75	20.40

***Source:** Own computation from Annual Reports of Banks*

From the two profitability measures ROE is volatile and ROA is relatively stable over the study period. Its value ranges between1 .32% and 4.42%. Similar to ROE, ROA has a declining trend after 2013. Banks registered the lowest profitability in 2002 as measured by both ROA and ROE. In particular, the average ROE of banks has plummeted sharply from 2000 to 2002. The lowest value of average ROE 4.21% registered in 2002. In contrast, the average ROE of banks increased sharply from 2002 to 2006 when it hit its highest value of 46.86%. From 2013 onwards the volatility of ROE has stabilized but declining.

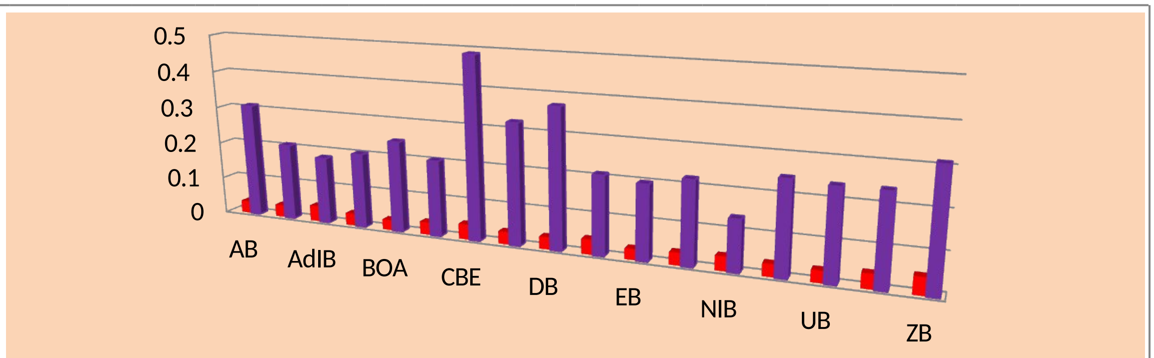
Figure 26: Trend of ROA and ROE of banks 1998 -2021



Source*: Own computation from annual reports of banks*

The bankwise profitability data shows that Commercial bank of Ethiopia, Dashen bank and Cooperative bank of Oromia, respectively, are the top three banks with the highest return on equity. On the other hand, Nib International bank, Addis International bank, and Enat bank are banks with the least return on equity. With regard to return on asset Zemen bank, Addis International bank, and Commercial bank of Ethiopia are the top three banks in the order they are presented. Bank of Abyssinia, Enat bank, and United bank have the lowest return on asset.

Figure 27: ROA and ROE of banks



	AB	AbB	AdIB	BIB	BOA	BuB	CBE	Coop	DB	GBE	EB	LB	NIB	OIB	UB	WB	ZB
ROA	3.38%	3.36%	4.32%	3.36%	2.84%	3.47%	4.15%	3.39%	3.33%	4.07%	2.93%	3.36%	3.83%	3.34%	3.21%	3.91%	4.64%
ROE	30.83%	20.61%	18.03%	20.09%	24.48%	20.33%	48.78%	32.27%	37.20%	20.90%	19.72%	21.97%	13.44%	24.41%	23.68%	23.76%	30.95%

Source*: Own computation from annual reports of banks*

4.2. Stochastic Frontier Analysis

4.2.1. Descriptive Statistics of Model Components

Table 5 below presents the standard descriptive statistics of the variables used in the stochastic frontier model. All the variables except time are log transformed. The „between standard deviation‟ measures the inter-bank variation while the „within variation‟ measures the variations along time. Except for loans and capital price the „within‟ variation is higher than the „between‟ standard deviation. This implies that the variation across time is relatively larger than the variation across banks.

Table 5: Descriptive statistics of model components

Variable		Mean	Std Dev.	Min	Max	Observations
Lntotal cost	Overall	19.41	1.27	16.05	23.27	N=264
	Between		0.84	18.09	21.77	n=17
	Within		0.91	16.44	21.17	T-bar 15.53
Lnlabor price	Overall	10.84	0.29	9.73	11.64	N=264
	Between		0.19	10.64	11.34	n=17
	Within		0.24	9.83	11.52	T-bar 15.53
LnFund price	Overall	-3.47	0.42	-5.47	-2.34	N=264
	Between		0.23	-3.91	-2.91	n=17
	Within		0.37	-5.38	-2.40	T-bar 15.53
Lncapital Price	Overall	-0.59	0.55	-3.57	0.54	N=264
	Between		0.36	-1.19	0.13	n=17
	Within		0.42	-3.14	0.80	T-bar 15.53
Lnloan	Overall	21.64	1.32	18.37	25.03	N=264
	Between		1.01	19.91	24.28	n=17
	Within		0.82	18.65	23.33	T-bar 15.53
LnEarning	Overall	20.34	2.09	13.29	25.54	N=264
	Between		1.25	18.98	24.34	n=17
	Within		1.55	13.06	22.64	T-bar 15.53
Lncapital	Overall	20.31	1.08	18.02	23.60	N=264
	Between		0.81	19.13	22.45	n=17
	Within		0.69	17.93	21.66	T-bar 15.53
LnNPL	Overall	-3.84	0.76	-6.18	-1.41	N=264
	Between		0.47	-4.68	-2.76	n=17
	Within		0.61	-6.36	-2.28	T-bar 15.53

Source: *Own computation from annual reports of banks*

4.2.2. Empirical Results

4.2.2.1. Estimation Results of the Stochastic Frontier Analysis

As stated in Chapter Three, the stochastic frontier analysis method is employed to measure the efficiency/inefficiency of banks. I have employed the Kumbhakar (2014) model that identifies both transient and persistent inefficiencies. Before taking the estimation results, I have run the Hauseman Test to choose between fixed effects and random effects estimation. The Hausman determines if errors are systematically related or not, that is to determine whether to apply the random effects or fixed effects model. The result of the Hausman test (χ^2, 71.42 and 0.0000) shows that the difference in the coefficients is systematic; hence the fixed effects model is more appropriate. As opposed to random effects, fixed effects allows the unobserved variables to have associations with the observed variables. Unlike random effects, the fixed effects estimate controls for the effects of time-invariant variables with time invariant effects. In effect, inefficiency is allowed to be correlated with the frontier regressors which avoids distributional assumptions.

Table 6 presents the empirical results from the regression analysis on the costs, outputs, and price of inputs using the Translog function. As indicated in Table 6, price of labor, loans (quadratic specification) ; capital (quadratic specification), the interaction between labor price and fund price, and the ratio of non-performing loans to total loans (both the linear and quadratic specifications) are positively associated with the cost of banks. In addition, the quadratic specification of other earning assets is positively associated with the cost of banks at 10% significance level. In contrast, loans and fund price (both the linear and quadratic specifications); the interaction of labor with time, with loans and with non-performing loans ratio; and the interaction of loan with non-performing loans ratio and with equity have negative coefficients.

The positive coefficient indicates a positive association between the variables and total cost of banks. For labor price a positive coefficient indicates that increases in the price of labor in banks is associated with increase in cost. Increase in labor costs per employee is directly reflected in the total cost of banks. Thence increase in labor price impacts bank efficiency negatively. This may mean that higher labor prices do not translate to higher productivity, and cost efficiency improvements. This is not in variance with the literature. Mosunda (2008) have found a positive labor price effect on the total cost of banks in Zambia. Conversely, an increase in the price of funds reduces the total cost of banks significantly. This could emanate from a reduction in mobilization costs of banks when deposit rates are higher or set through negotiation as in the case of fixed time deposits.

From the outputs, loans have a negative coefficient that is statistically significant at 1% significant level. Similarly, other earning assets have a negative, though insignificant, coefficient. This indicates a positive scale effect of loans. The negative coefficient of the interactions of loans with labor, with non-performing loans and with capital further indicates high scale effects of loans. There is no evidence on the scale effect of other earning assets.

With regards to non-performing loans, the result indicates a positive association between non-performing loans and total cost. That is, banks with higher non-performing loans ratio would have higher total costs than those with lower non-performing loans ratio. The impact of non- performing loans on the total costs of banks, thus on the efficiency of banks, is not straightforward because it depends on whether we have faced a bad luck , have a bad manager, or have make a strategic choice to skimp the underwriting and screening process (Mester and Deyoung, 1997). Under bad luck and bad management hypothesis, non-performing loans will have negative impact on the cost efficiency of banks. On the contrary, under skimping the increase in costs from non-performing loans, such as loan workout costs, default costs, and costs associated with recovering loans from collateral, are compensated by a reduction in underwriting and screening costs, at least in the short run. Even though the timing precedence of non- performing loans and cost efficiency of banks are not traced in this study, the positive coefficient of non-performing loans could imply the possibility of “Bad management” and/or “bad luck” hypothesis in the Ethiopian Banks.

The coefficient of equity captures the impact of a change in equity on bank costs in a particular year (Ivan et al, 2017). It is the measure of the shadow cost of equity. The shadow cost of equity is computed as the partial derivative of the cost function with respect to capital and shows the cost savings associated with an increase in the equity level of banks. The result shows that the coefficient of capital is positive and insignificant. This shows that a change in the level of capital does not affect the overall cost of banks. This is not surprising because bank creditors or depositors in Ethiopia do not reward banks with higher level of capital. However, the quadratic specification of capital has a positive and significant coefficient. In addition, the interaction of capital with loans has a positive coefficient that is significant at 5% level. On the other hand, the interaction of capital with non-performing loans has a negative coefficient that is significant at a 1% significant level.

Technical Progress

The coefficients of time in the regression output measure the changes in the total cost of banks across time. They indicate different types of technical progresses. Following Baltagi and Griffin (1988) and Heshmeti and Kumbhakar (1996) (both cited in Ivan et al, 2017) technical progress can be classified in to three: “pure-technological progress”, “scale-augmenting technological progress”, and “non-neutral technological progress”. The pure technical progress, measured by the time variable, both at the linear and quadratic specifications, is positive. The coefficient of „time‟ is 0.124 while the time-squared is -0.002; the latter is significant at a 5% level. Thus, there is no evidence of technical progress among banks in Ethiopia. In fact, although infinitesimal there seems to be a technological regress among banks in Ethiopia over the study period. The scale-augmenting technological progress is measured by taking the coefficients of the interaction term of time with outputs (Atlunbas et al., 1999). In this regard, the result indicates the coefficient of time with loans and other earning assets is negative but insignificant. Thus, there is no evidence of scale-augmenting technological progress too. Finally, the non-neutral technological progress is measured by taking the coefficients of time with input prices (Atlunbas et al. 1999). The non-neutral technological progress indicates a negative coefficient for labor indicating a declining share of labor costs in the total cost of banks. Conversely, albeit insignificant, the coefficient for financial cost is positive. Abdurezak and Tesfaye (2016) have also found no improvement in the efficiency of banks over the period 1997 to 2015.

The finding in this study is in agreement with Badunenko and Kumbhakar (2017) who have found technical regress among Indian banks. However, much of the extant literature evidences positive technical progress in banks over time; such as Atlunbas et al., (1999) and the Ivan et al (2019) among banks in Europe; Hunter and Timme (1991) among the large US banks; and Mosunda, albeit very small, have found positive technical progress.

Table 6: Estimation Results

Variable	Coefficient	Std. Error	T	$p > /t/$
Lnloan (Q_1)	- 2.169	1.205	2.47	0.014
LnEarning (Q_2)	-0.515	0.439	-1.17	0.241
Labor (P_1)	3.971	1.609	2.47	0.014
Lnfund(P_2)	-4.762	1.439	-3.31	0.001
LnCapital(E)	0.469	1.607	0.29	0.771
LnNPL	1.587	0.534	2.97	0.003
Time (T)	0.124	0.151	0.82	0.414
$(Q_1 Q_1)/2$	0.295	0.084	3.50	0.001
$(Q_2 Q_2)/2$	0.016	0.009	1.74	0.084
$(Q_1 Q_2)$	0.021	0.025	0.85	0.396
$(P_1 P_1)/2$	-0.049	0.115	-0.42	0.672
$(P_1 P_2)$	0.508	0.158	3.22	0.002
$(P_2 P_2)/2$	-0.346	0.080	-4.31	0.000
$(Q_1 P_1)$	-0.198	0.0855	-2.31	0.022
$(Q_1 P_2)$	0.0452	0.072	0.63	0.528
$(Q_2 P_1)$	0.014	0.028	0.49	0.622
$(Q_2 P_2)$	-0.017	0.199	-0.86	0.393
$(Q_1 E)$	-0.332	0.911	-3.64	0.000
$(Q_1 NPL)$	-0.147	0.046	-3.21	0.002
$(Q_2 E)$	-0.023	0.034	-0.68	0.500
$(Q_2 NPL)$	-0.007	0.014	-0.47	0.640
$(E E)/2$	0.388	0.121	3.21	0.002
$(NPL NPL)/2$	0.133	0.039	3.35	0.001
$(NPL E)$	0.188	0.060	3.11	0.002
$(P_1 E)$	0.049	0.118	0.41	0.680
$(P_1 NPL)$	-0.137	0.042	-3.26	0.001
$(P_2 E)$	0.061	0.096	0.63	0.529
$(P_2 NPL)$	-0.048	0.060	-0.80	0.422
$(T T)/2$	-0.002	0.001	2.16	0.032
$(T P_1)$	-0.026	0.008	-3.42	0.001
$(T P_2)$	0.004	0.006	0.67	0.503
$(T Q_1)$	-0.008	0.005	-1.64	0.102
$(T Q_2)$	-0.001	0.003	-0.50	0.620

Source: *Own computation from Annual Reports of Banks*

4.2.2.2. Efficiency Scores

The efficiency of banks is measured using Kumbhakar et al (2014) model. I have computed the overall transient and persistent inefficiencies/ efficiencies of banks from 1998 to 2021.

Persistent Inefficiency/efficiency

The result on the efficiency measures of the Ethiopian banks indicates that a significant portion of the banks‟ inefficiency is persistent inefficiency. This finding indicates that the efficiency of banks in Ethiopia is strongly dependent on structural factors such as location, client structure, macroeconomic environment, regulation, and operating model of the banks. The persistent inefficiency of banks amounts to 28.59%. The average persistent efficiency of banks is 76.78%. This suggests that an average bank uses 23.22% more resources than the bank at the efficiency frontier. Banks with the highest efficiency measures are Commercial Bank of Ethiopia, Dashen bank, and Awash bank; while banks with the lowest efficiencies are Enat bank, Addis International bank, and Global bank Ethiopia. Generally, larger and old banks seem to have higher persistent efficiency than younger and smaller banks.

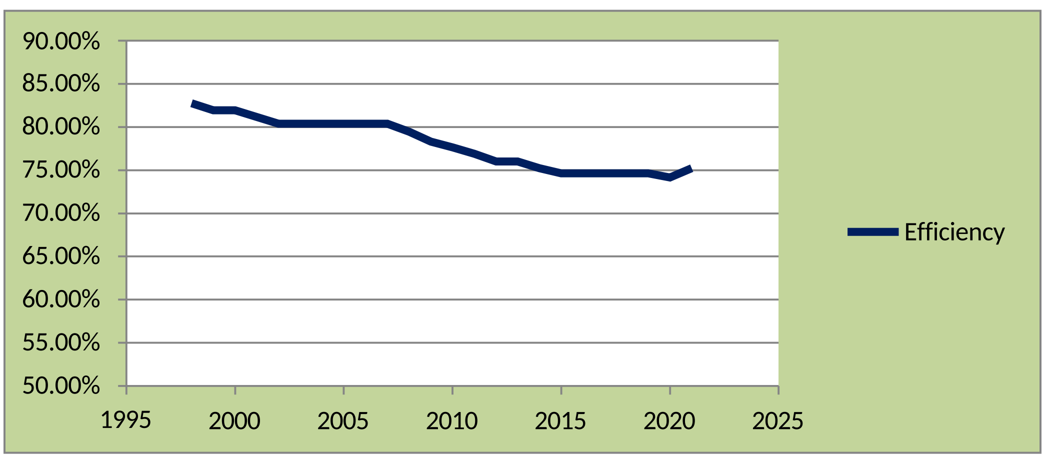
Transient Inefficiency

The average transient inefficiency measure of the banks from 1998 to 2021 is 2.32%. The transient efficiency of banks is 97.79%. This suggests that an average bank uses 2.21% more resources than the bank at the efficient frontier. The transient inefficiency of the banks is very low indicating that time-varying conditions are not the major source of the banks‟ inefficiency in Ethiopia. That is after controlling for bank heterogeneity and persistent efficiency, the share attributed to the time varying inefficiency is very small. This finding is in agreement to the study on European banks (Ivan et al., 2019). This implies that structural long-term factors seem to play bigger role for bank efficiency in Ethiopia than factors that change over time. The transient inefficiency of banks is very close to each other. The difference is infinitesimal. Regardless, the interbank comparison of the transient inefficiency among the banks indicates that Abay bank, Addis International bank, and Nib International bank are banks with the highest transient efficiency; whereas Zemen bank, Birhan International bank, and Oromia International bank are banks with the least transient efficiency. The result does not have a clear pattern in terms of the size or age of banks. Both small and large or old and young banks have the highest and lowest efficiency levels.

Overall Inefficiency

The overall efficiency of the banks is the product of persistent efficiency and transient efficiency. The overall efficiency of banks is 73.12%. That is an average bank employs 26.88% more resources than a bank at the frontier. I have also computed the overall efficiency of banks using another model, the Battesse Coelli (1995) time decay model. The result shows a comparable result of inefficiency, 26.08%. This result is in conformity with the 73% efficiency level found out by Rao and Lakew (2012). On the other hand, the result is lower than most of the results in previous studies such as 84% by Abdurezak and Tesfaye (2017); 92.5% by Emishaw (2016); 94.5% by Ram and Mesfin (2019) and 95.6% Gamachis (2016). The evidence on the efficiency level of banks anywhere else is diverse. There is a very high inefficiency level, as high as 50.7% by Bikker (2005). Yet other studies have found moderate and low inefficiency levels; as low as 6-9% (Mester, 1996). However, most studies found moderately high inefficiency measure comparable to this study such as Girardone et al., (2004) 30.2%; Ferrier and Lovell (1990) 26%; Atlunbas (1999) 20-25%. Banks with the highest efficiency measures are Commercial Bank of Ethiopia, Dashen bank, and Awash bank; while banks with the lowest efficiencies are Enat bank, Addis International bank, and Global bank Ethiopia. The banks with the highest efficiency levels are comparatively larger banks. Besides, for Commercial bank of Ethiopia, the source of efficiency could be the government ownership. It also serves as a treasury for most of the government institutions enabling it to generate deposits with out significant operating costs and enables it to generate much of its deposits from core deposits with lower costs. Commercial bank generates the lowest amount of its deposits, 3.18%, from non-core deposits.

Figure 28: Trend of overall efficiency 1998 - 2021



Source: *Own computation from Annual Reports of Banks*

Table 7: Persistent, transient and overall efficiency of banks

Bank	Persistent Efficiency	Transient Efficiency	Overall Efficiency
AB	81.18%	97.75%	79.35%
AbB	71.59%	98.76%	70.70%
AdIB	63.78%	98.66%	62.93%
BIB	72.89%	97.16%	70.82%
BOA	79.59%	97.46%	77.57%
BuB	71.36%	98.36%	70.19%
CBE	85.71%	97.26%	83.36%
Coop	78.89%	97.34%	76.79%
DB	81.94%	97.85%	80.18%
EB	63.21%	97.65%	61.72%
GBE	65.25%	98.26%	64.11%
LB	73.93%	98.11%	72.53%
NIB	76.77%	98.37%	75.52%
OIB	77.15%	97.25%	75.03%
UB	78.31%	97.61%	76.44%
WB	79.35%	97.33%	77.23%
ZB	70.58%	97.21%	68.61%
Total	76.78%	97.79%	73.12%

Source: *Own computation from Annual Reports of Banks*

4.2.2.3. Determinants of Inefficiency

The determinants of bank inefficiencies have been estimated using panel regression under the fixed effects model. The efficiency measures are regressed against expected bank efficiency determinants or correlates. The result for the persistent inefficiency values does not provide reportable results. Therefore, only the determinants of transient inefficiency are reported. Only bank level characteristics are considered. These variables are age, branch network, size (measured by total assets of a bank), non-core deposit ratio to deposits, non-core ratio to financial assets, earnings to deposit ratio, and intermediation ratios. Before taking the estimation results, I have conducted the Hausman test in order to determine if errors are systematically related or not and to determine whether to apply the random effects or fixed effects model. The result of the Hausman test (χ^2, 10.7 and 0.1524) shows that the difference in the coefficients is not systematic; hence the random effects model is more appropriate than the fixed effects model. Further the Lagrange Multiplier (LM) test to check whether the variances across entities is equal to zero which checks whether there is a panel effect or not. The test result indicates that there is panel effect thus results of the random effects estimate are used.

As indicated on the regression output in Table 7 below, the size of the impact of these variables on the value of transient inefficiency is infinetsmal. Turning to the relationships of the variables, ratio non core deposits to total deposits (*NonCore_ ratio*), ratio of non-core deposits to financial assets (*Noncoreratio_FA*) and intermediation ratio (*Intermediation_ratio*) have significant relationships. Non-core deposits to financial assets ratio (*Noncoreratio_FA*), total assets (*lnassets*) of banks and branch network (*lnbranches*) are negatively associated with the transient inefficiency of banks. The negative coefficients for total assets and branch networks, though insignificant, indicate that smaller banks are more flexible and suited to address time varying conditions that affect the efficiency of banks than the bigger ones. This findings is in conformity with Emishaw (2016) and Abdurezak and Tesfaye (2017).

On the other hand, the share of non-core deposits to total deposits (*NonCore_ ratio*), age (lnage) and intermediation ratio (*Intermediation_ratio*) have a positive coefficient that indicates a positive association with transient inefficiency. On the reverse, Emishaw (2017) has found negative correlation for intermediation ratio. The case of age on the effect of inefficiency could relate to the technological regress on the banks that imply increase in the cost of banks with an increase in time, *ceterus paribus*. The positive correlation between the intermediation ratio and transient inefficiency relates to, probably, a high screening and monitoring costs incurred to convert deposit to loans. These could imply the high cost incurred by banks to turn their deposits to productive outputs.

Table 8: Determinants of transient inefficiency

Variable	**Coefficient**	**Std. Error**	**Z**	$p > \|t\|$
Lnage	0.0000128	$5.86e^{-06}$	0.22	0.827
Lnbranches	$-1.91e^{-06}$	$-1.99e^{-06}$	-0.96	0.337
ROA	0.0001067	0.0001219	0.88	0.381
Noncore-Ratio	$5.05e^{-06}$	$4.34e^{-06}$	4.16	0.023
Noncoreratio_FA	-0.0000684	0.0000367	-1.86	0.062
LnAssets	$2.63e^{-07}$	$4.16e^{-06}$	0.06	0.950
Intermediation_ratio	0.000296	0.00011	2.69	0.005
-cons	0.0024036	0.0000806	29.82	0.000

Source: *Own computation from Annual Reports of Banks*

Chapter Five

Conclusion and Policy Implications

The financial system is a crucial part of the economy in a country. As a significant part of the financial system, banks intermediate funds from surplus units to those with fund deficits. Banks ameliorate the information asymmetry between borrowers and lenders, manage risks; channel funds to profitable investments, in effect, ensuring the smooth flow of resources within an economy. Inefficiency in the banking system makes the whole economic system inefficient and deters economic growth through high cost of intermediation. Therefore, problems faced by banks affect the whole economy substantially. For banks high efficiency reduces the chance of failure or ensures resilience. Numerous researches have evidenced that failing banks during the 1980s banking crisis are placed far from the best practice frontier (Berger and Humphrey, 1992; Berger and De Young, 1997).

As a primary financial institution in Ethiopia, the performance of banks affects the economy of the country enormously. Poor performance in the banking system begets sever implications for the overall economy. Currently, the banking system in Ethiopia is passing through major policy changes: several banks are joining the market and foreign banks are allowed to operate. This study evaluates cost efficiency of banks over the period 1998 to 2021 and identifies the determinants of cost efficiency among commercial banks in Ethiopia using the stochastic frontier analysis. The study employs the intermediation approach.

The result indicates that the overall efficiency of banks is 73.12%. That is an average bank employs 26.88% more resources than a bank at the frontier. The result shows that significant portion of the inefficiency among Ethiopian banks is persistent inefficiency. This indicates that efficiency of banks in Ethiopia is strongly dependent on structural factors. The transient inefficiency is insignificant, 2.32%, compared to the 28.59% for persistent inefficiency. The overall efficiency of banks is 73.12%. Generally, larger and older banks seem to have higher persistent efficiency than younger and smaller banks. With regard to pure technical progress, there is there is no evidence of technological progress among Ethiopian banks; in fact, even though very small, there is a regress in pure technical progress. In effect the efficiency of banks has a down ward trend over the years 1998 to 2021.

Regarding, the determinants of banks non-core deposits to financial assets ratio and total assets of banks are negatively associated with the transient inefficiency of banks. These indicate that smaller banks are more flexible and suited to address time varying conditions that affect the efficiency of banks than the bigger ones. On the other hand, the share of non-core deposits to total deposits, age, the ratio of other earning assets to deposits ratio, and intermediation ratio have a positive coefficient that indicates a positive association with transient inefficiency. The positive correlation between the intermediation ratio and transient inefficiency relates to, probably, a high screening and monitoring costs incurred to convert deposit to loans. The result is similar with the ratio of earning assets to deposits ratio. These could imply the high cost incurred by banks to turn their deposits to productive outputs.

Finally, the study also shows that increase in labor price affects bank efficiency negatively no productivity effect of labor price increases. Conversely, increase in fund price reduces the total cost of banks significantly mainly from reduction in mobilization costs of banks when deposit rates are set higher or through negotiation. Thus, deposit mobilization costs significantly influence the efficiency of commercial banks. From the outputs, loans are found to have a positive scale effect while there is no evidence on the scale effect of other earning assets. NPL has a negative impact on cost efficiency implying the possibility of "bad management" and/or bad luck" among the Ethiopian Banks.

In order to improve efficiency of commercial banks in Ethiopia bank managers, regulators and policy makers has to act on persistent inefficiency. In this case the focus has to be on factors that cause persistent inefficiency such as such as location, client structure, macroeconomic environment, regulation, and operating model of the banks. In addition, attention shall be given to improving the technical progress of banks through well designed skill developing training programs that can enhance the productivity and technical capability of the labor force.

Future research can extend bank efficiency through incorporating macroeconomic variable, comparing efficiency with other East-African and SSA; and the scope efficiency of banks through time. In addition, comparison of the results under different approaches, methods and input-output definitions can further assist to better understand cost efficiency among commercial banks in Ethiopia.

Bibliography

Abdurezak, M. & Tesfaye, B,. (2016) Cost Efficiency of Ethiopian Banks. *Ethiopian Journal of Business and Economics*, 6 (2). PP 126-158

Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* , 6, 21-37.

Aigner, D.J. and SF. Chu, (1968). On estimating the industry production function. *American Economic Review* Vol. (58) PP.826 - 839.

Akoena, S. K, Aboagye, A. Q., Antwi-Asare, T. O. and Gockel, F. A. (2009). „A study of Efficiencies of Ghanaian Banks‟. A paper presented at the CSAE conference 2009 on „Economic Development in Africa‟, University of Oxford, 22-24 March 2009.

Alemayehu G, 2006. The Structure and Performance of Ethiopia's Financial Sector in the Pre- & Post-Reform Period with a Special Focus on Banking, WIDER Working Paper Series RP2006-112, World Institute for Development Economic Research (UNU-WIDER).

Ally Z.,& Patel, D. G.(2013). Measuring Bank Efficiency:Evidence from the Tanzanian Banking Sector. *International journal of Management*

Alber, N., Elmofty, M., Kishk, I., & Sami, R. (2019). Banking Efficiency: Concepts, Drivers, Measures, Literature and Conceptual Model. DOI: http://dx.doi.org/10.2139/ssrn.3310982

Allen, Linda & Rai, Anoop, (1996). Operational efficiency in banking: An international comparison. *Journal of Banking & Finance, Elsevier.* Vol. 20(4), pages 655-672.

Athanasoglou, P., Brissimis, S. N., & Delis, M. D. (2008). Bank-Specific, Industry-Specific and Macroeconomic Determinants of Bank Profitability. *Journal of International Financial Markets, Institutions and Money*, 18, pages 121-136.

Altunbas, Y., Liu, M. H., Molyneux, P. and Seth, R. (2000). Efficiency and Risk in Japanese Banking. *Journal of Banking and Finance,* Vol. 24, pp. 1605–1628.

Altunbas, Y. & Gardener, E. P. M. & Molyneux, P. & Moore, B.,(2001). Efficiency in European banking. *European Economic Review, Elsevier*, Vol. 45(10), pages 1931-1955.

Badunenko, O & Kumbhakar, C.,(2017). Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research, Elsevier*, vol. 260(2), pages 789-803.

Battese, G., & Coelli, T. J. (1995). A model for Technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20, 325-332.

Beck, T. & Hesse, H, (2006). Bank efficiency, ownership, and market structure: why are interest spreads so high in Uganda?Policy Research Working Paper Series 4027, The World Bank.

Belotti, F., Daidone, S., Ilardi, G. and Atella, V. (2013) Stochastic frontier analysis using stata *The Stata Journal* 13(4), 719-758.

Benston, George J.(1972). "Economies of Scale of Financial Institutions. *Journal of Money, Credit, and Banking*, Vol. 4, 312-41.

Benston George J.,A. Hanweck and Humphrey D.B. (1982). Scale economies in banking: A restructuring and reassessment. *Journal of Money, Credit and Banking*, Vol. 14 (4), pp. 435-456. Ohio State University Press

Berger, A.N. and Humphrey, D.B. (1992) Measurement and Efficiency Issues in Commercial Banking. In: Griliches, Z., Ed., Output Measurement in the Service Sectors, National Bureau of Economic Research, Studies in Income and Wealth, Vol. 56, University of Chicago Press, Chicago, 24-79.

Berger, A.N. & Humphrey, D.B.(1997). Efficiency of financial institutions: International survey & directions for future research. *European Journal of Operational Research*, 98, 175-212.

Berger, A.N. and DeYoung R. (1997) Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance* 21, 849-870.

Berger, A. & Mester, L. (1997). Inside The Black Box: What Explains Differences in the Efficiency of Financial Institutions. *Journal of Banking and Finance*, 21, 895-947.

Bikker, J.A. (2004), Competition and Efficiency in a unified European banking market, Edward Elgar.

Bikker, J. A., & Bos, J. W. (2005). Competition and efficiency in banking: International comparisons. *Economic and Financial Modelling*,12(3), 103–148

Coelli T, Perelman S, Romano E (1999). Accounting for environmental influences in stochastic frontier models: with application to international airlines. *Journal of Productivity Analysis* 11:251–273. https://doi.org/10.1023/a:1007794121363

Coelli, T., Rao, D.,O‟Donnell, C. and Battese, G. (2005). An Introduction to Efficiency and Productivity Analysis, Springer, New York

Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014), Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-Run Efficiency, *Journal of Productivity Analysis*, 42(2), 123–36.

Cornwell, C., P. Schmidt, and Sickles R. C.(1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics* 46: 185–200

Drucker, P.F.(1963).Managing for business effectiveness. *Harvard Business Review,*1(3), 53–60.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica* 19, 273- 292.

Demirgüç-Kunt, A., Ross Levine and Hong G. (1998). Opening to Foreign Banks: Issues of Efficiency, Stability, and Growth. In Seongtae Lee (ed.) The Implications of Globalization of World Financial Markets. Seoul: The Bank of Korea.

Dinberu Y. and Wang., (2018) Drivers of Technical Efficiency of Ethiopian Commercial Banks: DEA Approach & Tobit Model. *Journal of Finance and Accounting*, Vol. 9(4) 30-38

Emishaw B. (2016) Determinants of Commercial Banks" Cost Efficiency in Ethiopia: A Stochastic Frontier Analysis, unpublished MSc Thesis, Addis Ababa University

Evanoff D, and Ors, D.(2008). The Competitive Dynamics of Geographic Deregulation in Banking: Implications for Productive Efficiency. *Journal of Money, Credit and Banking*, Vol.40, No. 5, pp.897-928.

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society*, 120(3), 253-281.

Ferrier, G., & Lovell, C. (1990). Measuring cost efficiency in banking: econometric and linear programming evidence. *Journal of Econometrics*, 46, 229-245.

Girardone, C., Molyneux, P. & Gardener, E. P. M. (2004). Analysing the determinants of bank efficiency: the case of Italian banks. *Applied Economics,* 36 (3), 215-227.

Harvey, Charles (1996). Banking Reform in Ethiopia, working paper. No. 37, Brighton.

Hauner, D. and Peiris S. (2005), Bank Efficiency and Competition in Low-Income Countries: The Case of Uganda", IMF Working Paper No. 05/240.

Huang T.H., & Wang M.H. (2002). Comparison of economic efficiency estimation methods: Parametric and non-parametric techniques. *The Manchester School*, 70(5), 682–709.

Hughes P. & Mester L, 2008. *Efficiency in banking: theory, practice, and evidence*, Working Papers 08-1, Federal Reserve Bank of Philadelphia.

Humphrey, D. B. (1990). Why Do Estimates of Bank Scale Economies Differ? *FRB Richmond Economic Review*, 76 (5), 38-50.

Ikapel, O., Gregory S, and Maurice M. (2023). “Determinants of Banking Sector Efficiency in Kenya: Application of Non-Parametric Data Envelopment Analysis (DEA) Model”. *Asian Journal of Economics, Business and Accounting* 23 (13):18-28.

Leibenstien, H., (1966) Allocative effciency vs. `X-effciency', *The American Economic Review*, 56, 392-415.

Ivan H., Reiner M., and Diego M., (2019). *The cost-efficiency and productivity growth of euro area banks,* No 2305, Working Paper Series, European Central Bank

Jondrow, J., C. A. K. Lovell, I. S. Materov, and Schmidt P.(1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19: 233–238

Kablan, S. (2010). *Banking Efficiency and Financial Development in Sub-Saharan Africa,* IMF Working paper 10/36 Washington: International Monetary Fund

Kirkipatrick, C., Murinde, V., & Tefula, M. (2008). The measurement and determinants of X efficiency in commercial banks in sub-Saharan Africa. *The European Journal of Finance*, 14 (7), 625–639

Kiyota, H. (2009). *Confronting the Global Financial Crisis: Bank Efficiency, Profitability and Banking System in Africa* [Paper Presentation], “Fostering Development in an Era of Financial and Economic Crises”, Addis Ababa, Ethiopia.

Koopmans, T.C., (1951) Analysis of Production as an Effcient Combination of Activities, in T.C.Koopmans(ed.), Activity Analysis of Production and Allocation, New York, Wiley

Kulik, J. (2017), “Technical efficiency and the methods of its measurement”, *Ekonometria*, Vol. 3 No. 57, pp. 72-87.

Kumbhakar, S.C. (1990), “Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency,” *Journal of Econometrics*, 46, 201–11

Kumbhakar, S. C., Lien, G. & Hardaker, J. B. (2014), Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis* 41(2), 321-337.

Kumbhakar, S. C., Wang, H. and Horncastle, A. P. (2015), *A practitioner's guide to stochastic frontier analysis using Stata*, Cambridge University Press.

Lang, G. & Welzel, P. (1996). Efficiency and technical progress in banking Empirical results for a panel of German cooperative banks, *Journal of Banking & Finance* vol. 20(6), pages 1003-1023.

Leibenstein H,. (1966) Allocative Efficiency vs. "X-Efficiency". *The American Economic Review* Vol. 56, No. 3, pp. 392-415

Lovell, C. A. K. (1993). Production Frontiers and Productive Efficiency.

Lovell, & S. S. Schmidt (1984), The Measurement of Productive Efficiency (pp. 3-67). Oxford University Press.

Maudos, J., J.M. Pastor, F. Perez and Quesada J. (2002): Cost and Profit Efficiency in European Banks, *Journal of International Financial Markets, Institutions & Money*, Vol. 12 (1), pp. 33–58.

McAllister, P.H. and McManus D., (1993), Resolving the scale efficiency puzzle in banking, *Journal of Banking and Finance*, 17, 389-405.

McKillop, D.G., Glass, J.C., and Morikawa, Y., (1996). The composite cost function and efficiency in giant Japanese banks. *Journal of Banking and Finance* 20, 1651-1671.

Mitchell, Karlyn & Onvural, Nur M, (1996). Economies of Scale and Scope at Large Commercial Banks: Evidence from the Fourier Flexible Functional Form, *Journal of Money, Credit and Banking*, Blackwell Publishing, vol. 28(2), pages 178-199

Meeusen, W. & van den Broeck, J. (1977).Efficiency estimation from Cobb-Douglas production functions with composed error term*, International Economic Review*,18, 435–44.

Ncube, M. (2011), *Efficiency of the Banking Sector in South Africa*, paper presented at the Fourth African Economic Conference 2009 on Fostering Development in an Era of Financial and Economic Crises, 11–13 November 2009, Addis Ababa , Ethiopia .

Mester, L. J. (1996). A study of bank efficiency taking into account risk-preferences. *Journal of Banking and Finance*, 20, 1025-1045.

Muluneh, A. (2006), *"Cost Efficiency of Commercial Banks in Ethiopia: An Application of A Stochastic Frontier Analysis''*, [Paper Presentation] "Fourth International Conference on the Ethiopian Economy ", Addis Ababa, Ethiopia.

Musonda, A. (2008), An Investigation into the Determinants of Cost Efficiency in the Zambian Banking Sector, [Paper Presentation] 13th Annual African Econometrics Society (AES) Conference, Pretoria, South Africa.

Mwega F., (2011). The Competitiveness and Efficiency of the Financial Services Sector in Africa: A Case Study of Kenya," African Development Review, African Development Bank, vol. 23(1), pages 44-59.

National Bank of Ethiopia (2017), The annual report of 2017‟, Addis Ababa, Ethiopia.

National Bank of Ethiopia (2013), The annual report of 2000-2013", Addis Ababa,Ethiopia.

Nguyen, BH, Sickles, RC, and Zelenyuk, V. (2021) *Efficiency Analysis with Stochastic Frontier Models Using Popular Statistical Softwares*. Working Paper Series No. WP09/2021. CEPA (Centre for Efficiency and Productivity Analysis), School of Economics, University of Queensland, Australia.

Pitt, M. M., and L.-F. Lee. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9: 43–64.

Ram B., & Mesfin E.(2019). Estimating Technical Efficiency of Selected Commercial Banks in Ethiopia Using Data Envelopment Analysis. *JETIR* Vol.6(3).

Rao, R. and Tekeste, B. (2012). Cost Efficiency and Ownership Structure of Commercial Banks in Ethiopia: An application of non-parametric approach. *European Journal of Business and Management.* 4(10), ISSN 2222-1905

Resti, A. (1997). Evaluating the Cost efficiency of the Italian Banking System: What can be learned from the joint application of parametric and non-parametric techniques. *Journal of banking and Finance*, 21, 221-250.

Samuel K. & Kamau A.,(2021). "Banking sector competition and intermediation efficiency in Kenya," African Development Review, African Development Bank, vol. 33(4), pages 648-661, December.

Schmidt P. & Sickles C. (1984). Production Frontiers and Panel Data. *Journal of Business & Economic Statistics* Vol. 2, No. 4 (Oct., 1984), pp. 367-374

Sealey, C.W., Jr. and J.T. Lindley, (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance* 32, 1251-1266.

Tadele, T. (2016). Determinants of Commercial Banks Efficiency: Evidence from Selected Commercial Banks of Ethiopia. *International Journal of Scientific and Research Publications*, 6(5), 551-556.

Ten Kate, Adriaan, Economic Efficiency as the Ultimate Goal of Competition Policy (2016). SSRN: https://ssrn.com/abstract=2740523 or http://dx.doi.org/10.2139/ssrn.2740523

Tefula, M. (2001) Evaluating the performance of commercial banks in Sub-Saharan Africa, unpublished PhD thesis, University of Birmingham.