Performance of Zambian Commercial Banks in the Post-Liberalisation Period

Evidence on Cost Efficiency, Competition and Market Power

A thesis submitted

by

Anthony Musonda Simpasa

For the award of

DOCTOR OF PHILOSOPHY

in the School of Economics

UNIVERSITY OF CAPE TOWN

May 2010

Thesis Supervisor: Professor Haim Abraham

Abstract

This study investigates three aspects important of performance for Zambia commercial banks. Specifically, the thesis addresses the aspect of cost efficiency and the factors that affect inefficiency performance. The study also empirically answers the policy question regarding the banks' exercise of market power and the low degree of competition. Using a richly assembled panel data set obtained from the Bank of Zambia on individual banks from 1998 to 2006, the thesis utilises theoretically sound methodologies in addressing these research questions.

The results from the analysis reveal the following. Firstly, using stochastic frontier estimation approach, cost inefficiency was estimated to be 8 percent. This means that mismanagement of resources was an impediment to the efficiency performance. Nonetheless, we observed a reduction in cost inefficiency over time, with domestic private banks displaying remarkable improvement. A combination of bank-specific and exogenous factors deterred banks from attaining optimal cost efficiency. Notably, impaired loans, asset concentration and macroeconomic instability undermined the banks' ability to operate optimally. Regulatory factors did not exacerbate cost inefficiency.

Secondly, Zambian banks operated in an oligopolistic set-up. Based on a methodology anchored in the New Empirical Industrial Organisation literature, the results of a competitive test showed that banks earned their revenue under conditions of monopolistic competition. This finding was buttressed by the estimated time varying Lerner Index, a measure of market power. The index showed that commercial banks set their prices above marginal cost by more than 50 percent. However, the degree of market power narrowed towards the end of the sample period. Market concentration, efficiency performance, diversity in revenue sources and regulatory intensity accounted for much of the banks' exercise of market power. On the other hand, the high proportion of interbank deposits, credit risk exposure and inflation dampened the banks' exercise of market power.

To our knowledge, this study is the first of its kind in Zambia. Therefore, the results of the thesis have important policy implications. More significantly, since there is room for deepening the degree of competition and furthering efficiency gains, regulatory authorities should strengthen measures aimed at ameliorating risk problems in the banking industry in a bid to lower the banks' exercise of market power. The authorities should also accelerate should also accelerate efforts of reducing recourse to Treasury bills as a deficit financing tool in order to negate the banks' appetite for securities as a source of revenue. This can be done by placing more emphasis on the legal and

institutional framework for resolving problem credit situations. This will intensify competition and propagate efficiency gains in the banking market. The authorities should also expeditiously tackle instability in the macroeconomic environment, particularly the high rate of inflation which hampered the banks’ revenue performance and exacerbated the exercise of market power.

Dedication

This dissertation is dedicated to my wife, Beatrice and our children – Matampulo, Kunda and Wanzya – for their strength and patience during times of difficulty.

Acknowledgment

This work received contributions from many people, too numerous to mention. However, first and foremost, the thesis owes its timely completion to Professor Haim Abraham-my supervisor who guided me throughout the process. I am immensely grateful to him for believing in me and the viability of my project and for his steadfast commitment and devotion to this cause. Special tribute is also paid to the African Economic Research Consortium (AERC) for funding my studies from inception to finish. Without the AERC's financial assistance, this thesis would not have seen the light of day. I am also highly indebted to the Bank of Zambia (BoZ)-my employer, for granting me paid study leave to pursue this dream. To Mwiza Mbewe, formerly of BoZ, I say thank you for tolerating my inquisitiveness on things that seemed obvious. I also wish to pay tribute to my lecturers and colleagues in the School of Economics at the University of Cape Town and at the Joint Facility for Electives (JFE) in Nairobi, Kenya for a warm camaraderie, and to the resource persons at the AERC's bi-annual thematic research workshops, for their input and direction at the early stages of the thesis. To my family, I say thank you for being a pillar in my life. The task you saw me begin is finally accomplished. Finally, but not the least, I am grateful to God my Creator, for sustaining me throughout the course of my studies.

List of Tables

List of Figures

Table of Contents

Chapter One

Introduction and Background

1.1 Introduction

Banks play a crucial role in the economy, shaped mainly by the nature of the functions they perform. These functions distinguish banks from other financial and non-financial firms. The first of these functions is that of intermediating between borrowers and lenders. In the course of undertaking financial intermediation, banks process different types of information pertaining to creditors and the ability of debtors to meet their obligations in a timely manner. The second and related function is acceptance and transformation of risks and providing liquidity in order to facilitate transactions. Liquidity creation and risk management and transformation are aimed at preventing financial instability emanating from bank runs (Diamond & Dybvig, 1983). Amelioration of information asymmetries and liquidity provision uniquely define bank managers' incentives in producing financial services more prudently, competitively and efficiently.

The role of information processing, liquidity creation and risk transformation is particularly relevant in view of the imperfections that characterise financial markets. Gurley and Shaw (1955) observed that financial intermediaries play an important role in alleviating the problem of trade frictions between borrowers and lenders. Inefficiencies in trade can arise when one or more of the transacting parties have superior information than the counterpart (van Damme, 1994). However, until the 1970s, these aspects were not incorporated into formal models of financial markets. Akerloff (1970) and Leland and Pyle (1977) were among the first to offer a formal analysis of information asymmetry in financial markets. Later works by Stiglitz and Weiss (1981), Santomero (1984), Dewatripont and Tirole (1994) and Freixas and Rochet (1997) showed that severity of market imperfections in financial intermediation can lead to sub-optimal outcomes.

Banks also enjoy economies of scale in transactions costs by pooling resources from many small investors and investing in bulk assets. Therefore, banks use their informational advantage to invest funds into profitable investment projects in order to minimise the opportunistic behaviour. Banks also engage in client screening to reduce the risk of default by borrowers. In so doing, they perform the function of delegated monitoring (Diamond, 1984; Allen & Santomero, 2001).

Furthermore, banks facilitate payments directly to the economy in the form of cheque clearing and wire transfer services. Since bank accounts are generally more attractive than holding cash, payments through the banking system are more preferred to cash payments. Therefore, any breakdowns in the banking system could produce gridlock to the payments system with harmful effects on the economy.

Finally, a competitive and efficiently functioning banking system is critical to the transmission of monetary policy to the rest of the economy. The highly liquid nature of bank demand deposits makes them readily acceptable by the public as the most widely used instrument of exchange in an economy. Given that deposits command a significant component of money supply, commercial banks play an important role in transmitting monetary policy actions to the real sector. Banks should be able to effectively undertake this task when operating in a more competitive and efficient environment. It is against this background that McKinnon (1973) and Shaw (1973) argued that a financial system that efficiently generates and allocates financial resources can contribute greatly to productivity and growth of the economy. Thus, informed by the McKinnon-Shaw hypothesis, many developing countries embarked upon the liberalisation of financial markets with a view to correct the imperfections and financial repression policies that characterised these economies. This hypothesis rests on the premise that liberalisation of financial markets ushers in market incentives that generally lead to efficient allocation of resources and improves performance of financial intermediaries, particularly commercial banks.

This thesis evaluates the performance of Zambian commercial banks during the post-liberalisation period. Specifically, the study focuses on cost efficiency, competition and market power during 1998-2006. Zambia undertook comprehensive economic reforms beginning in 1992, a significant component of which was financial liberalisation (see Appendix I for chronology of selected policy changes). Reform measures included liberalisation of interest rates; deregulation of movements of capital and elimination of portfolio restrictions, including foreign exchange transactions; easing of banking activity restrictions and strengthening of the regulatory and supervisory framework. In addition, the central bank adopted indirect instruments of monetary policy and streamlined its operations, focussing more on price stability as the core objective. Reforms also entailed restructuring the banking system to make it more responsive to the dictates of price and regulatory incentives.

After many years of interventionist policies, a major objective of financial liberalisation was the removal of operating obstacles in the banking sector in order to foster competition and efficiency.

This was in recognition that competition and efficiency would induce banks to provide the much needed funds for private sector investment to support rapid and sustainable economic growth.

1.2 Statement of the research problem

The opening up of the financial sector was followed up by a banking crisis between 1995 and 1998. However, since the end of the crisis, Zambian banks have posted significant gains, breaking the cycle of vulnerability that characterised the industry in the years following the implementation of sweeping financial reforms. The banking sector is now deemed relatively stable, with most banks boasting of health balance sheets and adequate levels of capitalisation. These gains have been credited to the improvement in the regulatory framework which was put in place after the banking crisis. Notwithstanding these gains, doubts about the level of competition and efficiency continue to linger in the minds of many. These concerns have been motivated by observed conduct and dominance of some commercial banks.

The Zambian banking system is one of the most concentrated in Sub-Saharan Africa (SSA). The three bank concentration ratio for assets, deposits and loans averages above 60 percent. Lending rates are also significantly higher than deposit rates and sticky downwards, resulting in very large interest rate spreads in excess of 20 percent. These spreads are also among the highest in Africa (IMF and World Bank, 2002). All these factors have been attributed to the dominance of few banks in the industry, which raises concern of possible collusive behaviour. The World Bank (2004) observed that the oligopolistic nature of the Zambian banking sector is one of the main causes of the rigidity in bank lending rates and wide interest spreads. According to economic theory, high levels of concentration and wide bank interest rate spreads may be evidence of the banks' exercise of market power and uncompetitive conduct in the industry. It may also suggest that banks operate inefficiently and therefore mask their inefficiencies through large interest margins.

The above factors, coupled with other structural problems continue to place the Zambian banking sector at the centre of policy debate. In particular, recent policy debate has focussed on improving accessibility to financial resources and making the banking sector an engine of economic transformation. Furthermore, policy makers continue to ponder on how best to stimulate competitiveness in the banking industry while balancing the act of ensuring safety and stability and prevention of unwarranted risk taking behaviour. Given that the banking industry has undergone structural transformation after the crisis, the increasing dominance of few large (foreign owned) banks is especially of major concern regarding the degree of contestability in the industry.[1]

1.3 Observations from industry assessment policy studies

In 2002, the International Monetary Fund (IMF) and the World Bank conducted an assessment of the financial sector under the auspices of the Financial Sector Assessment Programme (FSAP). Although this study did not empirically analyse the performance of the banking system, the report concluded that competition and efficiency among Zambian banks were low. The report noted further that the continued public ownership of the Zambia national commercial bank (ZNCB), one of the largest banks (by size of assets and branch network) particularly hindered competition in the industry. Accordingly, the report proposed that the bank be privatised in order to stimulate competition in the banking system. The assessment also noted that further prudential and regulatory reforms were necessary to remove bottlenecks that hindered banks from exploiting alternative ways of raising efficiency and increasing competitiveness. The reduction in government borrowing requirements was also seen to be particularly beneficial in bringing down interest rate margins which had curtailed financial intermediation.

The above factors were echoed in a Financial Sector Development Programme (FSDP) report launched by the Zambian government (Ministry of Finance and National Planning, 2004). This report was in response to the issues raised in the IMF/World Bank (2002) FSAP report. The FSDP is a comprehensive strategy to build and strengthen the financial sector infrastructure to enable it support economic diversification and sustainable growth. This financial sector blueprint suggested that dominance of foreign owned banks may be a possible source of collusive behaviour. High service fees and wide interest spreads were also seen as evidence of cost inefficiency responsible for the low level of financial intermediation and lack of innovation in the banking industry. The report also observed that the portfolio composition of banks' assets was little diversified as shown by a relatively high proportion of Treasury bills and foreign currency deposits held in off-shore accounts. Lending to the domestic private sector remained rudimentary, making the Zambian banking sector one of the least developed in sub-Saharan Africa.

[1] Generally, it is understood that a large number of relatively similar sized banks have better chances of competing amongst themselves than would otherwise be the case.

In 2006, the Zambian government commissioned the first in a series of studies to assess the impediments to accessing financial services. This study, conducted by FinMark Trust (2006) looked at the demand side of the market. Based on survey evidence, the main observation was that the Zambian banking sector was characterised by high transactions costs which were passed on to customers in form of fees and other service charges. The study attributed the low level of financial accessibility to these factors, among others.[2] The second study implemented by Oxford Policy Management (2007) in conjunction with FinMark Trust, focussed on the supply side. The conclusion from this second survey was that the key constraints to financial accessibility were mainly bank-specific rather than of macroeconomic or regulatory nature. They were inherent in the internal organisational and structure of the banking market characterised by the cycle of heavy dependence on fees, net liquidity and high spreads which had resulted in very high costs for bank users.

All the issues discussed above were also raised in de Luna Martinez's (2006) study of the accessibility to financial services in the Zambian banking sector. Using descriptive analysis, the author observed that despite the policies of open doors to foreign bank participation, existing foreign owned banks have not been able to widen the range of bank products offered to both households and firms. This may suggest that there is inadequate level of competition and innovation in the industry.

According to Aryeetey, et al. (1994) these shortcomings broadly reflect deep-rooted information and other structural problems characteristic of undeveloped financial systems. The implication being that financial liberalisation measures may have ignored such structural factors as branching and accessibility to financial services, loyal customer base and creation of a market niche by incumbents, which dissuade new entrants and other fringe banks from gaining a competitive position. Thus, due to the inherent nature of banking systems in developing countries, policy reforms may not yield a 'critical level' of competitive pressure in the aftermath of deregulation as envisaged.

This perceived failure of reforms may also be due to the rapid pace at which the policies were implemented and the inability to recognise that each economy embodies intrinsic features which uniquely define its response to policy options. For example, in the banking industry, incumbent banks with an already entrenched position tend to have little motivation to embrace measures aimed at improving their cost efficiency levels (Sarkar & Bhaumik, 1998). Combined, these forces may not bring optimal outcomes in the banking sector and may in fact result in a financial crisis, as did

[2] Other impediments were the generally low levels of incomes among Zambians and the concentration of banks in urban centres, leaving a large part of the Zambian population in rural districts unbanked.

happen in Zambia. Indeed, the benefits of removing institutional and structural impediments to fostering banking competition and enhancing efficiency are well established (Northcott, 2004).

1.4 Justification of the research

Commercial banks in Zambia are the most active players in the financial sector, ahead of other financial sector participants.[3] However, the amount of credit dispensed to the private sector is low and this is evidently clear from the low accessibility to financial services by a large segment of the population. At 8 percent, the ratio of banks' private sector credit to the Gross Domestic Product (GDP) is one of the lowest in Sub-Saharan Africa. Zambian banks provide a relatively similar amount of credit to the public sector. This has tended to keep interest rates at high levels, resulting in financial disintermediation (crowding out of the private sector). In this regard, Zambian banks have had a limited impact on the economy in terms of private credit provision.

The assessment of the Zambian banking industry by the IMF and World Bank (2002), FinMark Trust (2006), de Luna Martinez (2006) and others, demonstrates a genuine concern of the predicament facing the Zambian banking sector vis-à-vis competition and efficiency. For bank practitioners, competition and operational efficiency are important parameters of evaluating performance in a dynamically changing environment. These changes affect the behaviour of individual banks to unimaginable proportions. Therefore, bank managers should comprehend these changes in order to gain an insight into the potential effect on the level of performance relative to their counterparts. Specifically, bank managers need to bring strategic planning in their organisation in order to enhance cost effectiveness and aim at delivering high quality banking products and services to their customers.

For the authorities, concerns about inadequate competitive conditions and low efficiency should serve as indicators of a poorly performing banking sector with adverse implications on service delivery and social welfare. Without doubt, knowledge of commercial banks' performance helps in understanding the role banks play in the national economy. However, without knowledge of indicators backed by theoretically plausible and empirically tested methodologies, the policy analysis and response would be lopsided. Therefore, research on the performance of banks is a relevant area

[3] As at end 2008, commercial banks had more than ten times the value of assets relative to other financial sector players put together.

of empirical enquiry especially for a low income country faced with numerous social challenges and perception about failed financial reforms.

This thesis is the first attempt at such an undertaking. The main theme of the thesis is to empirically respond to the policy concerns highlighted in the preceding sections. Given the existing research void in this area, the Zambian banking sector presents a strong case for the empirical analysis of its performance. This is achieved by appealing to tested methodologies based on sound theoretical microeconomic foundations.[4] Results from this exercise should guide the authorities in better understanding the dynamics affecting the banking industry in Zambia. Accordingly, the study is both timely and relevant to the present policy concerns.

1.5 Objectives of the research

The objectives of this study are shaped around the three empirical chapters and stem from the title of the thesis, that is, to evaluate the performance of the Zambian banking system. First, as observed in the preceding analysis and as will be shown further below, Zambian banks are alleged to operate inefficiently. Key indicators of cost inefficiency include high operating costs and poor delivery of financial services. Second, the high level of concentration and dominance of large banks and state ownership of some banks have been blamed for the inadequate level of competition. Whereas these indicators may be used to analyse competition, they fall short of adequately capturing banks' actual behaviour. Related to this, to the extent that few banks dominate the sector, collusive pricing may be the natural outcome. Therefore, the third aspect of the study relates to the banks' exercise of market power. This is empirically investigated using pricing and cost data. In sum, the study aims at investigating three pillars of bank performance, vis-à-vis cost inefficiency, competition and related market structure, and evidence of market power. These are important factors in devising anti-trust policy.

In answering these research questions we also seek to evaluate the degree to which macroeconomic and regulatory factors influence the banks' behaviour. For example, to the extent

[4] In the FSAP report, the IMF/World Bank regress spreads on other factors to measure causes of high interest rate spreads while Mbewe (2004) uses descriptive analysis to undertaken this task. A study by Mwilwa (2007) was the more comprehensive analysis of determinants of interest rate spreads, but even this took an aggregate view. As has been argued elsewhere, this approach is inadequate as a measure of competition and efficiency in banking, see World Bank, (2006), among others.

that regulatory policy does not promote competition in the banking industry, banks' exercise of market power could be sustained over a longer period. When regulatory policy stimulates competition among banks, this could drive banks' price-cost margins down, thereby eroding their market power. The price-cost margin, typically defined as price less marginal cost divided by price or the Lerner Index, has a long history of theoretical presence and empirical applications in industrial organisation. We also analyse the role of internal bank specific factors in perpetuating inefficiency and engendering market power in the banking industry.

Specific objectives, methods of analysis and tested hypotheses are articulated in each of the three empirical chapters investigating the elements of bank performance highlighted above.

1.6 Structure of the thesis

Chapter One has introduced the main theme of the thesis by providing a description of the statement of the problem, outlining the research questions and rationale of the study. In addition, the chapter has highlighted observations made by previous assessments of the Zambian banking sector. These issues serve as a basis for the general direction of the thesis.

Chapter Two puts the study in a policy and historical context, highlighting the nature of the Zambian banking industry and reforms that occurred to usher in the current financial landscape. Chapter two also brings together different performance indicators in the Zambian banking industry. These measures form a basis for the empirical analysis in later chapters.

Chapter Three is a presentation of some of the microeconomic theories/models of the banking firm. This chapter is critical to the understanding of the analytical approach taken in the empirical sections because it highlights the importance of these theories to the analysis of efficiency performance and competitive conduct in banking.

Chapter Four is the beginning of empirical analysis. Specifically, chapter four addresses the concept of cost efficiency in Zambian banking and how this is affected by banks' own internal factors as well as the regulatory and macroeconomic environment in which they operate. The basis for this analysis is underscored by the high level of operating costs and wide spreads observed in the banking industry. To be able to get a better understanding of banks' performance, a solid analytical framework is needed and this chapter brings out this aspect very clearly. In particular, the investigation of cost efficiency and its determinants is undertaken in the context of the single step methodology proposed by Battese and Coelli (1995).

Chapter Five begins by highlighting the importance of competition in the banking sector. Accordingly, in this part of the thesis, we investigate the state of competition in the Zambian banking sector, drawing from previous research and utilising tools of analysis based on the New Empirical Industrial Organisation (NEIO) literature. Specifically we address the problem by applying the methodology proposed by Rosse and Panzar (1977) and Panzar and Rosse (1982; 1987). The investigation is motivated by concerns of the high level of concentration in the banking sector and claims that subsidiaries of multinational banks which dominate the industry pose a threat to overall competitiveness and development of the Zambian banking sector.

Chapter Six extends the concept of competition by analysing the banks' exercise of market power using the time varying Lerner Index estimated from a translog cost function. Since the Lerner Index is estimated for each bank across time, we assess the evolution of market power and examine in detail the possible factors that influence the banks' conduct.

Chapter Seven is a concluding chapter of the thesis and brings together a summary of results from chapters four, five and six. This chapter also presents policy recommendations arising from the analyses. This is important for the authorities, bank practitioners and market analysts alike to make informed decisions.

Chapter Two

Financial Policies and Peformance of the Banking Sector

2.1 Introduction

The Zambian banking sector has emerged from the constraints imposed by financial repression policies to the current open policy with no explicit entry barriers, save for the mandatory minimum paid up capital for setting up a bank. In the following sections, we conduct an analysis of the performance of the banking system during the interventionist regime and in the post-reform period. This assessment is important because it puts the empirical analysis developed in subsequent chapters into context and enables us to appreciate the achievements and shortcomings in the level of banks' performance.

2.2 Financial repression policies and the banking Sector

After independence in 1964, Zambia adopted the import substitution industrialisation strategy as a development strategy. To achieve this objective, the government announced sweeping measures in 1968 and 1969, which included nationalisation of financial institutions, except foreign commercial banks despite their dominance in the industry (Harvey, 1973; Brownbridge, 1998).[5] Instead, foreign banks were incorporated in the country and operated as subsidiaries of foreign banks under the Zambian banking law which stipulated that half the directors of these banks should be resident in the country. However, the lending behaviour of foreign owned banks largely favoured the expatriate community and foreign owned companies.

Therefore, to avail funds and serve the interests of the indigenous businesses and parastatal companies not reached by existing foreign owned commercial banks, the government established state banks (Brownbridge, 1998). In addition, the government also created agricultural and development banks to provide long-term credit to 'strategic' sectors of the economy such as mining, and agricultural industries. To further provide impetus to the development agenda, the government

[5] Commercial banks were not nationalised because the expatriates threatened to withdraw their management staff and the Zambian government was not confident that the banks would perform without them.

also imposed controls on interest and exchange rates, transactions in and allocation of foreign exchange and applied stringent reserve requirements as tools of credit control. In addition, banks were required to seek permission from the central bank before they could extend credit to foreign companies or engage in bank-to-bank refinancing. In this regard, state enterprises were especially given preferential treatment in accessing foreign exchange while foreign companies were mandated to surrender foreign exchange to the central bank. This meant that the central bank maintained monopoly on the allocation of foreign exchange which was used as a credit rationing tool.

The authorities also imposed minimum paid up requirements for both domestic and foreign banks. However, these requirements were never revised; making them worthless in view of high rates of inflation. For example, by 1991 before reforms were undertaken, soaring inflation had reduced the real value of the minimum capital requirement to around US$0.3 million. At the time reforms were being initiated in 1992, this figure had been eroded to just more than US$50,000 in real terms. The reduction in paid up capital severely undermined safety and soundness of the financial system (Brownbridge, 1998).

During the interventionist period, regulatory rules governing the operations of the banking system were archaic and outdated leading to delinquent lending practices. The poor state of banking supervision and prudential regulatory framework was cited as the main cause of severe undercapitalisation of the banking sector. For many banks, the capital-to-assets ratio fell below the internationally recommended threshold. To aggravate the situation, interest rate controls meant that banks could not adequately make allowance for loan maturity or risk. Accordingly, prudence was undermined and by the early 1980s, banks had accumulated huge non-performing assets precipitated by non-repayment of loans by state enterprises, mainly because most investments were made in commercially unsuccessful ventures.

Unlike public banks, foreign owned banks successfully resisted political pressure to support unviable projects and their lending practices were guided by strict commercial criteria (Brownbridge, 1998). Therefore, the low quality of assets was not a major problem. Nonetheless, due to limited international competition arising from tight restrictions on current and capital account transactions, their scope of operations remained rudimentary and undiversified.[6]

[6] All foreign capital flows and investments were tightly regulated and subject to approval by the authorities.

Lack of data on the microeconomic performance of the banking sector during the interventionist period hampers an informed assessment of the performance of the banking sector under a financially repressive regime.[7] However, the above narrative suggests that government's involvement in the financial system coupled with numerous administrative controls had a deleterious effect on the performance of the banking system and on the quality and quantity of the banks' assets. It also undermined financial expansion and intermediation because its policies produced different forms of distortions which negated efficient functioning of a larger section of the economy. For example, ceilings on interest rates, subsidies and directed lending intended to promote economic growth constrained the expansion of credit thereby starving the economy of much needed development finance. When lending occurred, banks charged interest rates that were not commensurate with perceived risk. The subsidies and other distorting incentives granted to the public sector banks only helped perpetuate uncompetitive behaviour in the sector.

At a macro level, available statistics indicate that the spread between the lending and deposit rate averaged 5 percent between 1980 and 1989 while both these rates were negative in real terms due to a sustained high level of inflation. As a measure of financial depth, broad money to GDP ratio was less than 30 percent on average over the same period. On the other hand, credit to government as a share of GDP averaged 12.4 percent over the same period against 12.0 percent lending to the private sector, also expressed as a proportion of GDP. These numbers suggest financial disintermediation which was induced by financial repression policies and crowding out effects of government borrowing.

2.3 Financial liberalisation and the changing landscape of the Zambian banking industry

Financial sector reforms in Zambia commenced in 1992 after the change of government and were implemented in the context of a broader adjustment and stabilisation programme under the auspices of the International Monetary Fund (IMF) and the World Bank.[8] Generally, the reforms

---

[7]

[8] Faced with mounting opposition in the late 1980s as a result of increasing political autocracy and a discredited economic strategy, President Kaunda whose United National Independence Party (UNIP) ruled Zambia for nearly three decades, yielded to pressures for the restoration of multiparty democratic politics and voluntarily stood for re-election in 1991. However, he was defeated by Frederick Chiluba of the Movement for Multiparty Democracy (MMD) which heralded the beginning of major political and economic reforms.

progressed unhindered in the early years, spurred mainly by the need to redress the structural imbalances that characterised the previous regime. The major policy changes included liberalisation of interest rates, removal of restrictions on foreign exchange transactions and allocation mechanisms, easing of banking restrictions and improvement in the regulatory and supervisory framework with concomitant efforts to improve the efficiency of the existing banks. The deregulation of interest rates and removal of restrictions on credit advances were aimed at exerting increased pressure on banks to adopt more competitive practices in the pricing of bank products while allowing them to cut operating costs in order to stem inefficiency.

The reforms also entailed restructuring of the banking system by emphasising recapitalisation of problem banks in order to improve the state of the financial sector. Thus, changes were made to the regulatory framework while amendments to the Banking and Financial Services Act (BFSA) and the Bank of Zambia Act were also undertaken. These changes granted the central bank greater authority over banks and nonbanks. Licensing of commercial banks also shifted from the Ministry of Finance to the central bank, thereby consolidating the functions of the latter and giving it greater degree of oversight. These amendments helped realign regulation by function rather than by type of bank as was previously the case when public sector banks received preferential regulatory treatment compared to those imposed on foreign owned banks. In line with international best practices, the Bank of Zambia as the regulatory authority introduced a set of reporting requirements for commercial banks and guidelines for asset valuation and loan-loss provisions.

The reforms also redefined the central bank's responsibilities, focussing principally on price stability rather than on multiple roles as was the case in the previous system of control. Complementary to these policies was the lifting of restrictions on capital flows following the liberalisation of the capital account in 1994. The authorities also reformed the reserve requirements regime and allowed banks to operate foreign currency accounts in the same year. A year later, the central bank granted licences to the bureaux de change, allowing them to trade in foreign currency.[9]

Prior to 1992, there were only three local private banks existing alongside foreign and public sector banks. But these banks operated on the fringe of the market, and therefore accounted for a minute share of the industry assets. As shown above, the banking sector was subjected to a myriad of

[9] Subsequent to this, the foreign exchnage market has undergone significant reforms, culminating in the implementation of the broad-based interbank foreign exchange market (IFEM) system in July 2003, which undoubtedly has introduced transparency in the allocation and trading of foreign exchange.

distortionary policies, including restrictions on branching and cumbersome licensing requirements, which prevented entry into the sector. For this reason, incumbent private foreign and government owned banks took up much of the market share, accounting for over 70.0 percent of total assets. The removal of implicit and explicit entry restrictions into the banking industry beginning from 1992 saw an increase in branching activity by both foreign and public sector banks.

More importantly, the liberal operating environment saw a rapid increase in the number of local private banks, completely altering the landscape of the Zambian banking industry. Specifically, between 1992 and 1996, 12 new commercial banks were registered, bringing the total to 26 banks. By mid-1996, there were 13 local private sector banks operating in Zambia. It must be noted that the majority of these banks sprung up before reforms to the regulatory and prudential system were implemented. For instance, financial entry requirements remained low due to high inflation (Brownbridge, 1998). Therefore the low real paid up capital served as an entry stimulus for these banks. The period between 1992 and 1996 was also marked by a credit boom, with private sector credit as share of GDP increasing from 4.7 percent to 8.8 percent.

Nonetheless, the majority of these banks engaged in foreign exchange transactions and government securities trading. According to Brownbridge (1998) the extent to which the local banks relied on Treasury bills (Tbs) for their income was indicated by the high proportion of Tbs in total assets. Just in two years (1992-1993), Treasury bill holdings accounted for 28 percent of total assets while the share of bank credit was only 15 percent. The FSAP report by the IMF/World Bank (2002) makes similar observations that foreign exchange trading has traditionally been the main source of non-interest revenue for Zambian banks while revenue from Treasury bills accounts for a substantial amount of interest income.

In view of these conditions, local private banks took up enormous amounts of deposits while lending was mainly to the small traders which had been shut out from credit by large foreign owned banks.[10] Insider borrowing presented another problem. This exposed many of the banks to credit and exchange risk, exacerbated by instability in the macroeconomic environment characterised by high rate of inflation, depreciating currency and deteriorating external terms of trade. As a result, they

[10] Although the entry of new local private banks introduced some level of competition on the deposit side of the market (Brownbridge & Gayi, 1999), the ratio of total deposits to GDP did not increase proportionately, averaging only 13 % between 1989 and 1995. More than anything else, this reflected a shift of deposits away from large banks to new small ones due to an aggressive marketing strategy and attractive terms they were offering.

started to experience severe liquidity problems due to an increase in the level of impaired assets relative to short-term liabilities. Within a short period, some of these banks became insolvent, resulting in wholesale liquidations between 1994 and 1998. During this period, nine local banks were closed (see Table 2-1 below). The economic cost of these bank failures has been estimated at approximately 7.0 percent of GDP (de Luna Martinez, 2006).

Table 2-1: Types of banks in Zambia

Name of bank (Year established)	Type of ownership	Failed local banks[b]
Barclays Bank (1918)	Foreign private	Chase Trust Bank (1995-1997)
Standard Chartered Bank (1906)	Foreign private	Credit Africa Bank (1994-1998)
Stanbic Bank (1956)	Foreign private	First Merchant Bank (1994-1998)
Citigroup (1979)	Foreign private	Manifold Investment Bank (1987-1997)
African banking corporation (2000)	Foreign private	Prudence Bank (1994-1997)
Intermarket Banking Corporation (1995)	Foreign private	Meridien BIAO Bank (1984-1995)
Zambia National Commercial Bank (1968)	Local Public	Union Bank (1979-2000)
Indo-Zambia Bank (1984)	Joint (defined as foreign public)	Commerce Bank (1984-2001)
Bank of China (1996)	Foreign public	United Bank of Zambia (1997-2001)
Finance Bank (1986)	Local private	African Commercial Bank (1984-1995)
Investrust Bank Plc (1996)	Local private	Safe Deposit Bank (19923-1996)
First Alliance Bank (1994)	Local private	Mercantile Bank (1993-1995)
New Capital Bank (1989)	Local private	Continental Bank (1993-1995)
Cavmont Merchant Bank (1993)	Local private	Ital Bank (1993-1996)
Cavmont Capital Bank (2004)[a]	Local private	

Source: Bank of Zambia [http://www.boz.zm]; IMF/World Bank (2002) and Maimbo (2002)
[a]Merger between New Capital Bank and Cavmont Merchant Bank
[b]Period of operating in brackets.

The banking crisis has been blamed on a weak regulatory framework and an understaffed supervisory authority which caused banks to engage in imprudent lending behaviour (Maimbo, 2002; IMF and World Bank, 2002). Thus, in an attempt to prevent future crises, the central bank responded by strengthening its prudential regulatory and supervisory capacity, culminating in a restructuring programme of other ailing but solvent banks. Accordingly, since 1998, only three banks have been closed down, two because of failure to recapitalise and the third had its operating licence revoked in 2001 due to imprudent banking conduct related to money laundering activities. At the end of 2006, there were 13 commercial banks operating in Zambia. These comprised six subsidiaries of foreign banks, five domestic private banks, and three public sector banks. Public sector banks included one wholly owned by the Zambian government and two other foreign public banks.[11]

2.4 Selected performance indicators in the banking sector

This section presents indicators of performance for the Zambian commercial banking industry. These indicators pertain to major asset items, branching activity, market concentration, cost structure, profitability and aggregate risk.

2.4.1 Major assets of Zambian banks

The asset composition of all commercial banks shows that the share of loans and advances to total assets increased to 36.8 percent in 2006 from 21.0 percent recorded in 2002. Treasury securities accounted for a further 20.5 percent in 2006, down from 22.7 percent in 2002. The growth in the proportion of loans indicated a rebound in credit extension which coincided with favourable macroeconomic environment characterised by high economic growth and a slowdown in the rate of inflation. It also manifested a shift from securities holdings as yields became less attractive. This is evident in a reduction of Treasury bill holdings by 2.2 percentage points. Thus, during 2003 to 2006, total loans and holding of government securities accounted for just over half (51.0 percent) of the total assets. This figure reflected an improvement from 45.7 percent recorded between 1998 and 2002.

The remaining 49.0 percent of assets was divided as follows. Foreign currency deposits held abroad remained the largest component at 17.5 percent of total assets during the same period. However, this amount decreased from an average of 30.0 percent recorded between 1998 and 2002. The reduction in the proportion of foreign assets came against a backdrop of rapid appreciation of the domestic currency, which downgraded the Kwacha equivalent of these assets.[12] Another major asset

[11] The public sector banks comprised the Zambia National Commercial Bank (ZNBC) wholly owned by the state but has been privatised and transfer of ownership was concluded in April 2007, a Chinese state bank and a joint venture between the governments of Zambia and India.

[12] It is instructive to note that in dollar terms, foreign currency deposits grew by 30.6 % from 1998/2002 to 2002/2006 periods. Therefore, the observed reduction in domestic currency equivalent was as a result of currency appreciation.

item was the reserves held at the central bank. These accounted for 12.4 percent between 1998 and 2002 but grew by 2.6 percentage points to 14.9 percent, as a share of total assets. Interbank activity was a tiny fraction at 2.4 percent, a reduction from 4.0 percent recorded during 1998 and 2002. Similar amounts were recorded for fixed assets and other assets, respectively. In a nutshell, the analysis shows that the banks' earnings assets were in excess of two thirds of total assets. Banks' reserves at the central bank do not attract interest payment.

2.4.2 Branching activity

Table 2-2 below summarises the evolution of the structure of the banking market, vis-à-vis branching activity and level of concentration for assets, loans and deposits. The table shows that since the end of the banking crisis in 1998, the total number of banks changed very little. In 1998, there were 13 banks operating a total of 150 branches across the country. Of the 13 banks, seven were foreign owned (two of which were public foreign owned), five local private banks and one public domestic bank. The public domestic bank had the largest branch network of 42 branches spread throughout the country, including the rural areas. Foreign owned banks operated a total of 67 branches, nearly half of which were operated by Barclays bank, the largest bank in Zambia.

Table 2-2: Structure of the Zambian commercial banking sector, post crisis period

	1998			2002			2006		
	Foreign	Local	Public	Foreign	Local	Public	Foreign	Local	Public
Number of banks	7	5	1	8	5	1	8	4	1
Number of Branches	67	41	42	69	41	43	60	58	43
Total assets (K'bn)	900.39	139.07	373.47	3,199.76	443.70	979.16	7,260.49	1,575.26	1,676.72
percent of industry assets	63.73	9.84	26.43	67.73	9.39	20.73	68.01	14.76	15.71
Loans (K'bn)	288.94	28.65	142.18	825.85	83.29	65.61	2,739.78	609.33	517.53
percent of industry loans	62.84	6.23	30.92	84.73	8.54	6.73	70.86	15.76	13.38
Securities (K'bn)	109.75	18.58	26.07	528.27	135.37	383.81	1,354.23	274.54	521.50
percent of total industry holdings	71.08	12.03	16.88	50.43	12.92	36.64	62.98	12.77	24.25
Deposits (K'bn)	675.49	77.65	240.71	2,139.14	188.21	771.70	5,293.52	1,133.81	1,337.66
percent of industry deposits	67.97	7.81	24.22	69.03	6.07	24.90	68.17	14.60	17.23

Source: Bank of Zambia and own calculations

The number of operating banks increased slightly to 14 in 2002, owing to the licensing of a new foreign bank, the African Banking Corporation (ABC) in 2000, bringing the total number of foreign banks to eight. Other categories of banks remained unchanged until 2004, when a merger between New Capital Bank and Cavmont Merchant Bank was authorised, forming Cavmont Capital Bank (CCB) at the beginning of 2004. This is reflected in the reduced number of operating banks to 13 in 2006. Together, these 13 banks operated a total of 161 branches in 2006, against 150 in 1998, representing growth in branching activity of 7.3 percent. The increase in branching activity arose from a countrywide expansion in the number of branches by Finance bank, one of the most successful local private banks. Specifically, it acquired some branches originally operated by Barclays bank, and opened a few new ones. The closure of some branches by Barclays bank led to a 15.0 percent reduction in the total number of branches operated by foreign owned banks.[13]

2.4.3 Market growth and concentration

Table 2.2 also shows the growth in banking activity as well as the breakdown in the degree of concentration. From Table 2.2, aggregate industry assets grew substantially over the period, amounting to K10.7 trillion (US$2.9 billion) at the end of 2006. A third of this amount was held by foreign owned banks, depicting an 8.2 percent increase over the 1998 share. In value, the total amount of assets held by foreign banks more than doubled to K7.3 trillion (approximately US$2.0 billion). The remainder was shared by local private banks and the public bank, which collectively accounted for 30.5 percent of industry assets, representing K3.3 billion (approximately US$0.9 billion) in value.

The dominance of foreign owned banks as measured by the value of assets predates the crisis and is a reflection of an entrenched position in all segments of the banking market. For the credit market, loans held by foreign banks amounted to K2.7 trillion (US$0.7 billion) in 2006, a growth of 231 percent over 2002 position in nominal terms. The surge in advances came against a backdrop of renewed lending by most foreign banks, following an improvement in the macroeconomic environment since 2002. Again, during the same period, domestic banks and the state-owned bank lagged behind, recording total loans worth K1.1 trillion (US$0.3 billion) between them. As a

[13] Barclays bank closed almost half of its branches due to dwindling level of profitability of these branch offices in view of the changing market conditions induced by falling Treasury bill fortunes and rapid appreciation of the domestic currency, both of which caused a diminution of earnings. However, it would seem that Barclays bank management seriously miscalculated the ecnomic fundamentals because within a year of closing these branches, there was a in policy shift, which saw an even more aggressive branching activity and recruitment of staff to man the new offices.

proportion of total industry loans, private domestic banks accounted for 15.8 percent and the public sector bank took up 13.4 percent. Measured against the 2002 level, both categories of banks recorded an improvement, even though the dominance of foreign banks remained evident despite shrinkage in their share of loans.

On the liability side, purchased funds accounted for the largest share, taking up 80.5 percent of total liabilities. Accordingly, the following analysis pertains to the deposit side of the market. Here again we observe a dominating position by foreign banks throughout the review period. The stock of deposits held by foreign banks stood at K5.3 trillion (US$1.4 billion) in 2006, two thirds of the industry total deposits for the period. This amount also depicted a growth rate of 147.5 percent in nominal value over the 2002 position. Controlling for inflation effects, the growth rate in deposits for foreign banks was only 32.1 percent; that for domestic private banks was substantially higher at 221.5 percent while the value of real deposits held by the public bank shrunk by 7.5 percent (increased by 73.3 percent in nominal terms).

The above indicators depict the relative dominant position of foreign banks in the Zambian banking industry, a concern raised by the authorities as being inimical to overall competition in the sector. Table 2.3 shows more clearly the level of concentration in the banking sector in relation to the three-bank concentration ratio $(CR3)$ for assets, deposits and loans, respectively.

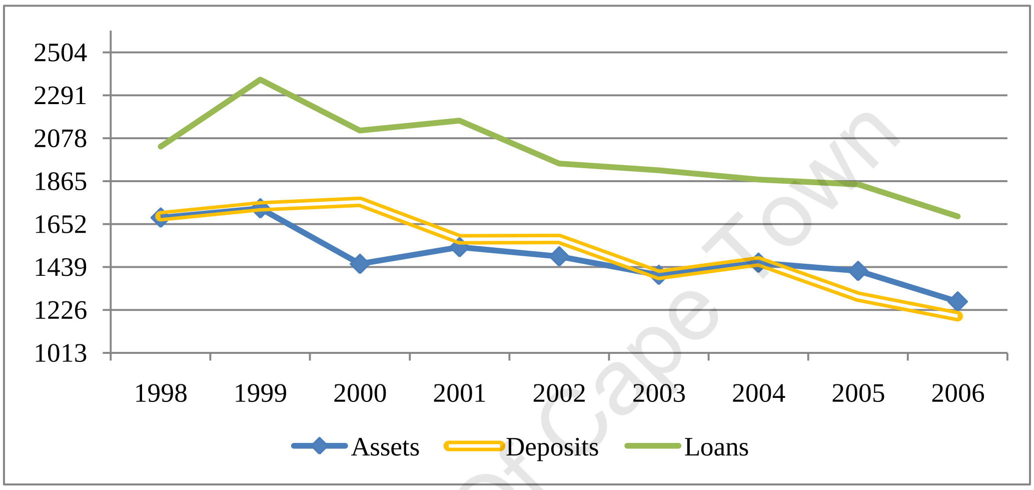
Table 2-3: Concentration in the Zambian commercial banking sector

Three bank concentration ratios, CR3 (percent):1998-2006			
	Assets	Deposits	Loans
1998	64.3	67.7	71.2
1999	66.2	69.3	77.1
2000	59.4	61.9	71.5
2001	60.3	62.9	74.0
2002	59.2	62.4	62.7
2003	55.8	58.5	60.6
2004	58.1	61.5	61.6
2005	55.8	56.7	60.1
2006	50.4	53.1	59.7
Average	58.4	61.6	66.5

Source: BoZ data and author’s own calculations

It can be seen from Table 2.3 that the three largest banks accounted for 58 percent of assets, 67 percent of loans and 62 percent of deposits. Of the three banks included in the calculation of the ratio, two are foreign owned banks and the third is a public bank, which accounted for more than a fifth in each of the industry assets, deposits and loans. Figure 2.1 depicts the Herfindahl Hirschman Index (HHI) which also shows a similar pattern.[14]

Figure 2.1: Banking concentration - Herfindahl-Hirschman Index (HHI)



Source: BoZ data and author's own calculations

From the chart we observe that the level of concentration was high for the loans market since the *HHI* exceeds the threshold of 1800 which regulators use to measure intensity of competition, indicating clear dominance of a few banks in this segment of the market. Yet again, foreign banks account for the largest proportion of the estimated *HHI*. The *HHI* for deposits and total assets shows a moderate level of concentration because it falls between 1000 and 1800. Both $CR3$ and the HHI show that despite a slight downward trend, the banking sector is still relatively more concentrated, depicting some evidence of market power. However, given the pitfalls of these structural measures in evaluating competitive conduct (World Bank, 2006; Hausman & Sidak, 2007), these conclusions

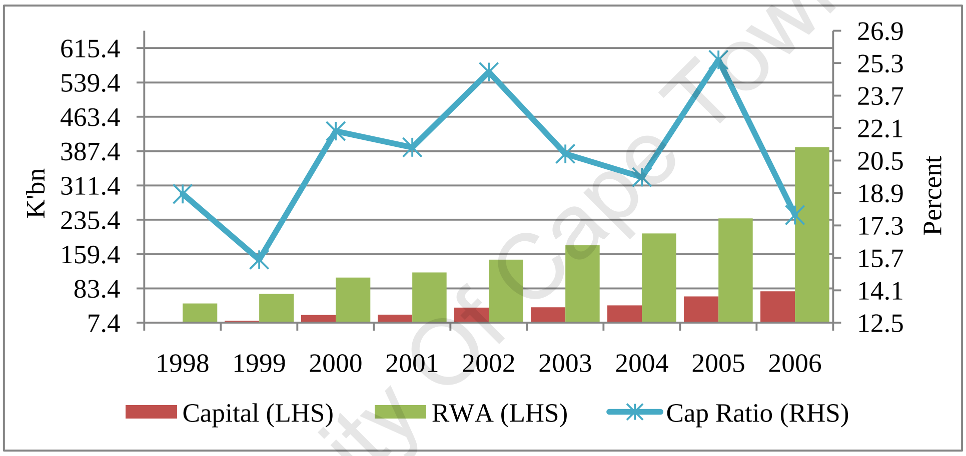
[14] The HHI is calculated as $HHI = \sum_{i=1}^{N} s_i^2$ where s_i^2 is the square of a bank's market share.

may be flawed even though the *HHI* is more preferred to the *CR3* as a measure of market structure because of its sensitivity to the number of firms and relative firm distribution (Amato, 1995).

2.4.4 Bank capitalisation and solvency

Since the end of the banking crisis, solvency concerns have lessened as commemrcial banks have generally recorded satisfactory capital performance (see Figure 2.2).

Figure 2.2: Aggregate capital adequacy position



Source: BoZ data and author's own computations

Figure 2.2 shows that the aggregate capital adequacy ratio (Cap Ratio) defined as total regulatory capital-to-risk weighted assets (RWA) has been increasing, but was at its lowest level in 1999. Coming from a crisis, this year represents troubling times for small local banks. However, quick action by the regulatory authorities averted further deterioration in capital through the enactment of the Bank of Zambia Act of 1996 and the amendment in 2000 of the Banking and Finacial Services Act (BFSA) of 1994.[15] Against this background, Zambian banks continued to exhibit an increasing level of compliance with the capital adequacy requirements, negating major concerns of solvency risk.

[15] The amendment of the BFSA in 2000 enhanced the effectiveness of supervision of non-bank financial institutions.

In 2006, Tier I capital as a proportion of risk-weighted assets averaged 20.7 percent, well above the 8 percent recommended under the Basle II Accord. It was also in excess of the regulator's own threshold of 10 percent. A similar picture is observed for total regulatory capital to risk-weighted assets, which amounted to 22.4 percent over the sample period. Thus the relative stability in the banking sector since the turbulent early years of reform has been ascribed to the improvement in regulatory and supervision capacity of the central bank. The existing pieces of legislation are now seen to provide sufficient scope for the central bank to preemptively deal with any problem banks.

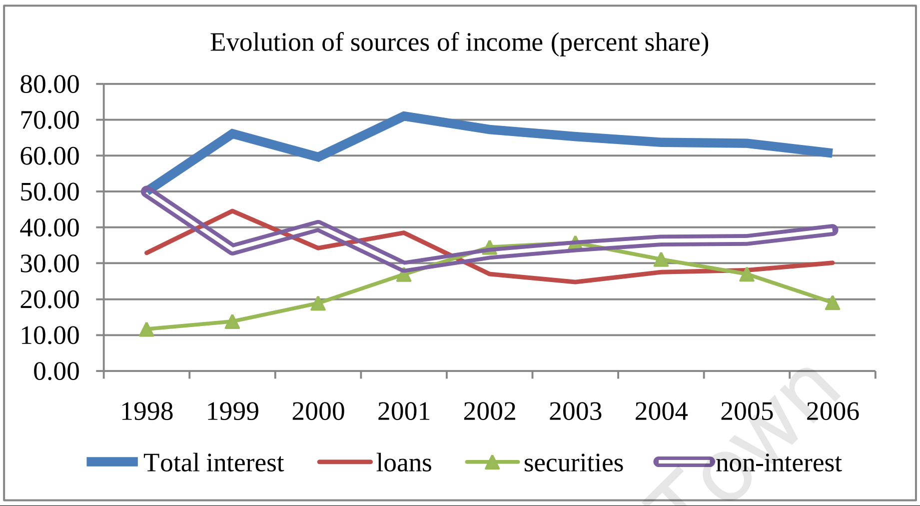
In relative terms, foreign owned banks boasted of better capitalisation than domestic banks, although both categories of banks remained compliant with the minimum requirements. The ratio of Tier I capital-to-risk weighted assets for foreign owned banks stood at an average of 23 percent between 1998 and 2006 against 17 percent for domestic private banks for the sample period. Evaluated against total regulatory capital, foreign banks had a capital-risk weighted assets ratio of 24 percent compared with 19 percent for local banks.

From the above, it is no surprise that the Zambian banking system has been described as adequately capitalised and in a better shape than before the crisis (IMF and World Bank, 2002; Ministry of Finance and National Planning, 2004). The IMF/World Bank FSAP report actually gave Zambian banks a bill of approval, stating that despite other potential risks in the sector, these risks did not threaten banks' solvency given high capitalisation ratios.

2.4.5 Sources of bank revenue

Figure 2.3 below shows the evolution of three main sources of income, as a proportion of total operating income. Zambian banks earn their revenue from three main sources: interest on advances; interest on investment in securities and non-interest income (earned from fees, commissions and gains on foreign exchange transactions). For the industry as a whole, the main source of revenue was interest income, averaging above 60 percent of total revenue. Over the years, interest income on loans (including net interest income from interbank placements) has been the largest component of total interest revenue, exceeding that of treasury securities. However, the share of loans interest income has been thinning over the years due to the obliterating effect of non-performing loans and accompanying provisioning. Nonetheless, between 2001 and 2004 the reduction in loans interest revenue was offset by an increase in interest earned from investments in Government securities which bolstered the banks' revenue position.

Figure 2.3: Major sources of banks' income (percent share of total income)



Source: Author's own calculations based on BoZ data

In 2002, income from securities exceeded revenue from loan advances, thereby narrowing the gap between these two revenue sources. This trend continued until 2004 when the government implemented a domestic debt consolidation programme, culminating in significant reduction in borrowing requirements which underpinned the disinflation process. This programme led to a sharp decline in yield rates on securities, causing a concomitant decline in interest income on these investments. As a result, the proportion of revenue earned from securities shrunk and with this, a rebound in the share of loans interest revenue.

In light of the above, it can be seen that interest income from securities as share of total income amounted to only 11.6 percent in 1998, but increased threefold to 34.5 percent in 2002, before it tumbled in 2004 to 31.1 percent, coinciding with the fall in yield rates. As the fall in yield rates steepened in 2006, the share of securities income in total revenue shrunk to its 2000 level of 19 percent. The reduction in government borrowing requirement stemmed from an improvement in the macroeconomic profile, which also served as a catalyst for increased private sector lending by the banks, beginning in 2004. Indeed, de Luna Martinez (2006) reports that Barclays bank, the largest lender in the industry only resumed making loans to the private sector after a drastic fall in Treasury bill yield rates. Although the share of loans interest earnings in total interest income declined progressively from a peak of 32.9 percent in 1998 to 27 percent in 2002 it rebounded in 2004 to 27.6 percent and increased even further to 30.1 percent in 2006 though it remained below the 1998 level.

However, the increase in the rate of non-performing loans from 2003/4 significantly dampened the uptake in loans income. Consequently, loans interest revenue grew less than the fall in returns on securities, implying a higher decrease in total interest income through to 2006. This is also highlighted in Figure 2.3 above.

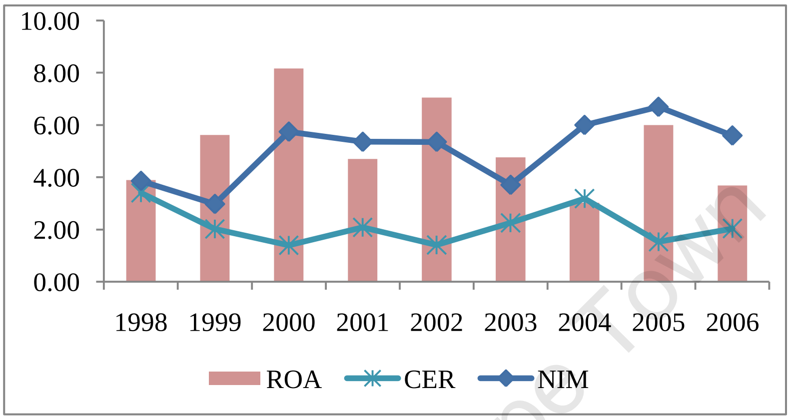
The steady increase in income from securities also acted as disincentive for exploiting other sources of income, outside traditional intermediation activities. In particular, since the banking crisis which saw a curb on speculative foreign exchange transactions, non-interest revenue declined, mirroring an expansion in traditional sources of revenue. Therefore, income on fees, commissions and gains from foreign exchange transactions exhibited volatility till 2001, just before the government instituted measures to put a further clamp on dollarisation activities. Subsequently, non-interest income flattened somewhat, falling below the 40 percent mark as proportion of total income. The IMF/World Bank (2002) also made similar observations in the FSAP that foreign exchange gains suffered a knock in 2001 due to the relative stability of the Kwacha during that year.

Although the share of non-interest income became a less important source of revenue between 2001 and 2003, it has since recouped the sparkle of the early years, driven largely by diminishing income from securities. Thus, between 2004 and 2006, non-interest income accounted for an average of 37.4 percent against 34.6 percent recorded from 2001 to 2003. From 1998 to 2000, the proportion of non-interest income to total operating revenue amounted to 38.3 percent. The resurgence in non-interest income was mainly attributed to fees and commissions, which offset the reduction in gains from foreign exchange trading as the domestic currency appreciated steeply from the third quarter of 2005. Between 1998 and 2004 fee income as a proportion of non-interest income amounted to 36 percent while gains from foreign exchange transactions accounted for 52 percent. The appreciation in the domestic currency saw a shift in the relative importance of these revenue sources. Accordingly, between 2005 and 2006, banks posted 47 percent in fees and commissions and 45 percent in gains from foreign exchange trading.

2.4.6 Cost structure and profitability of commercial banks

Strong earnings and profitability reflect a bank's capacity to build an adequate capital base necessary to absorb losses, finance its expansion and meet its obligations to shareholders. The widely used indicator of profitability is the return on assets (ROA) which is often supplemented by the return on equity (ROE). The net interest margin (NIM) is another common measure of bank profitability. Figure 2.4 below shows the evolution of banks' aggregate costs and profitability indicators over the sample period. Zambian banks boast of high profits, underpinned by the three main income streams discussed above.

Figure 2.4: Aggregate profitability and cost-earnings ratio



Source: Bank of Zambia, and author's own estimates

The pattern of profits as measured by the return on assets (ROA) mirrored the evolution in net interest margin (NIM) defined as net interest revenue divided by earnings assets. Volatility characterised the movements in both ROA and NIM, especially between 2002 and 2006. The ROA ranged from a low of 3.0 percent in 2004 to a high of 6.0 percent recorded in 2006. Similarly for NIM, it fell to a low of 3.0 percent in 2003 and reached a peak of 6.7 percent in 2005, the highest since 1998. However, both indicators reflected movements in the banks' cost structure depicted by the cost-earnings ratio (CER) that is, ratio of operating costs to net income.

Between 1998 and 2000, the level of profitability was strong, allowing banks to record an annual average of 5.9 percent return on assets. This was also consistent with the increase in net interest margin (NIM) across the industry. However, both ROA and NIM dipped in 2001, as banks lost significant amount of revenue due to exchange losses. The decrease in ROA was greater than that of NIM, mainly because interest earnings remained high relative to interest expenses.

Several factors explain the level of profitability. Most importantly, they reflect a sustained upward trend in yield rates for government securities more than a reduction in operating costs. Therefore, without a major shift in revenue sources, banks remain vulnerable to sharp declines in government borrowing requirements and the accompanying shrinkage in yield rates on Treasury securities. This was especially evident during 2005-2006 when yield rates on Treasury securities plummeted to record lows, falling below 10 percent since 1998. Accordingly, both ROA and NIM declined.

Another explanation for the wide interest margins may be macroeconomic instability as given by the high and variable rate of inflation. Over the sample period, inflation averaged 20.5 percent mainly because of government's appetite for deficit financing during the 1990s and early 2000s. Between 1998 and 2002, the rate of inflation averaged 24.8 percent but fell drastically to around 16 percent between 2003 and 2006, thanks to stringent fiscal operations complemented by a strong anti-inflationary monetary policy pursued by the Bank of Zambia. These factors eventually culminated in the drastic fall in the inflation rate to 8.2 percent at the end of 2006. This was the first time the rate of inflation had hit single digit in three decades.

Until 2007, commercial banks had suffered high and unremunerated reserve requirements. In 2003, the reserve ratio on liquid assets was reduced to 14 percent from 17.5 percent to induce lending to the private sector. However, banks responded rather sluggishly, prompting the central bank to lower the ratio further to 8.0 percent in 2007. Nonetheless, long years of high reserve requirements served as a disincentive for meaningful reduction in lending rates, which remained persistently higher than deposit rates, leading to a wide interest rate spread and net interest margin. At a disaggregated level, Table 2.4 below shows profitability by bank ownership structure. Foreign owned banks recorded higher profits than domestic private and public sector banks.

Table 2-4: Bank profitability measures, by ownership category

	Return on assets, ROA (percent)				Net interest margin, NIM (percent)			
	Foreign	Local	Public	All	Foreign	Local	Public	All
1998	4.80	0.85	2.75	3.89	5.47	2.29	0.39	3.86
1999	7.92	0.47	0.17	5.61	4.53	1.01	1.81	2.97
2000	9.99	3.12	4.32	8.16	6.72	4.17	1.66	5.74
2001	7.48	5.93	-7.16	4.70	5.65	4.81	1.82	5.35
2002	7.15	7.60	6.43	7.05	4.64	5.68	5.29	5.35
2003	5.65	6.27	0.93	4.76	2.85	5.80	0.09	3.70
2004	3.14	4.94	1.31	3.02	5.10	7.54	5.01	6.00
2005	6.52	6.99	3.15	6.00	5.25	7.71	9.15	6.70
2006	3.98	5.27	1.33	3.68	4.75	4.84	7.18	5.60
Average	6.29	4.61	1.47	5.21	5.00	4.87	3.60	5.03

Source: Calculated from BoZ data

At an average of 6.3 percent, foreign owned banks' level of profits was 37 percent higher than that for domestic private banks and was 4.3 times greater than that posted by the public sector banks. A striking feature of the banks' profitability is that foreign owned banks lost their first position to domestic private banks between 2002 and 2006, thanks to the latter group's rationalisation of operational costs.[16] During this period, domestic banks performed better in profitability terms.

From 2002-2006, the return on assets and net interest margin for domestic banks averaged 6.2 percent and 6.3 percent, respectively. For foreign banks, profitability ratios were 5.3 percent for ROA and 4.5 percent for the NIM over the same period. In the preceding three years (1998-2001), domestic private banks recorded ROA of 2.6 percent on average and NIM averaging 3.1 percent against 7.5 percent and 5.6 percent, respectively for their foreign owned counterparts. The improvement in profit performance showed that domestic private banks had recouped the loss in profitability since the banking crisis which dampened their earning opportunities as a result of flight to quality.

The performance of the state owned bank was dogged by a sharp increase in losses, which saw negative profits in 2001. However, in 2002 profits rebounded strongly but this momentum was short-lived as profits dissipated, reeling below 2 percent on average between 2003 and 2006. Therefore, the recovery of 2002 was insufficient to wipe out the earlier losses mainly due to a reduction in fee income and a stable currency which substantially eroded foreign exchange gains. This underscores the bank's vulnerability to foreign exchange trading to cushion its income against falling interest earnings.

From the data above, except during 2002-2006, the state owned bank had the lowest level of NIM. Prior to this period, the public sector bank lagged behind all bank categories in NIM, posting only 1.4 percent between 1998 and 2001. However, a clean up in the balance sheet boosted the quality of assets and with this, a revamp in earnings and profitability. Accordingly, between 2002 and 2006, the public sector bank recorded NIM of 5.4 percent, even higher than that for foreign owned banks.

[16] The expense ratio (i.e., ratio of non-interest expense to net operating revenue) amounted to 59.0 percent for domestic private banks excluding the state owned bank during 2002-2006 while that for foreign banks was 70.0 percent over the same period. However, from 1998-2001, domestic private banks' expense ratio was 0.22 percentage points higher than foreign owned banks. Clearly therefore, domestic banks instituted cost saving measures in the latter part of the study period, which resulted into better profit performance.

The wide margins observed from the preceding discussion are a reflection of the oligopolistic nature and conduct in the Zambian banking market. With non-interest expenses exceeding three quarters of total operating costs, this suggests that Zambian banks have scope to capture efficiency gains by reducing operating costs, especially those related to employee compensation and benefits.

2.4.7 Measures of financial development and intermediation in the Zambian banking sector

The Zambian banking industry is also underdeveloped and exhibits low intermediation (see Table 2.5 below. Relative to Gross Domestic Product (GDP), total banking assets amounted to 25.3 percent between 1998 and 2001 while during 2002-2006, this figure increased marginally to 26.8 percent. Equally, reforms notwithstanding, the ratio of private sector credit to GDP amounted to only 7.6 percent between 2002 and 2006, a marginal increase from 7.1 percent recorded during the 1998-2001 period.

Table 2-5:Indicators of financial development and intermediation (Percent of GDP)

	ASSETS	PUBLIC CREDIT	PRIVATE CREDIT
1998	23.44	2.54	6.52
1999	24.63	3.10	6.98
2000	26.10	3.54	8.20
2001	26.90	5.92	6.87
2002	29.05	5.64	5.89
2003	26.79	8.76	6.57
2004	27.13	6.93	7.92
2005	23.84	7.31	7.23
2006	27.22	6.05	10.16
1998-2001	**25.27**	**3.77**	**7.14**
2002-2006	**26.81**	**6.94**	**7.55**
1998-2006	**26.12**	**5.53**	**7.37**

Source: IMF (IFS online edition); BoZ (Macroeconomic indicators) and own calculations

The banks' private credit as a share of GDP is one of the lowest in Sub-Saharan Africa and was marginally higher than credit to the public sector. Between 2002 and 2006, the annual average ratio of private credit to GDP was 7.6 percent, a slight improvement over the preceding period from 1998-2001. Over the corresponding period, the ratio of credit to government averaged 6.9 percent per annum, an expansion of 84.1 percent relative to the period from 1998 to 2001. The bulk of bank credit to government was accumulated during 2002-2004 before fiscal consolidation took effect. Subsequently, as noted above, debt accumulation receded culminating in low growth of fresh issuance of Treasury securities. Nonetheless, the relatively high proportion of credit to the public sector reflected preference for the low risky securities over more risky private sector lending. This also reflects the underdeveloped nature of the Zambian financial system.

2.4.8 Implications of observations in chapter two

Unlike in developed countries and emerging markets, the Zambian banking industry has not experienced any form of consolidation, whether market driven or government induced. Rather, the observed levels of industry concentration stem mainly from the dominance of foreign owned banks, which pre-dates independence. Although their dominance has been reduced somewhat, this is too small to erode their market share.

The literature shows that foreign bank presence in the domestic banking sector is good for competition and efficiency (Claessens & Laeven, 2004; Gelos & Roldos, 2004; Jayaratne & Stratan, 1998). These benefits are amplified by the process of financial liberalisation. However, in the case of Zambia, de Luna Martinez (2006) argues otherwise, suggesting that despite policies of open doors, competition and efficiency are low and there is no product innovation among foreign banks. The banking industry is also characterised by high operating costs, wide margins and profitability ratios. The level of financial intermediation is also at low levels.

This is in stark contrast with the theoretical prediction of the McKinnon-Shaw hypothesis that financial liberalisation fosters financial intermediation by encouraging savings mobilisation and thereby promoting financial development. In the Zambian banking industry, savings are constrained by negative real deposit interest rates while credit to the private sector is hampered by the high cost of lending and other structural impediments, including prohibitive loan pre-qualification requirements and the banks' appetite for relatively less risk government debt instruments. The relatively high proportion of foreign currency assets in total assets has perpetuated the financial disintermediation because banks enjoy better returns on foreign currency deneominated assets than on domestic private sector lending.

Whether the observed high level of concentration is evidence of efficient operations as propagated by the efficient-structure hypothesis (Demsetz, 1973; Peltzman, 1977) or a general tendency towards non-competitive behaviour is unclear. The exact form of competitive conduct and degree of efficiency can only be determined by interacting measures of market structure and bank behaviour. Without such an assessment, we cannot predict with any degree of certainty about the implications of this concentration and related high profitability ratios.

To the banks' credit, the level of solvency and liquidity are sufficiently high indicating low probability of systemic risks in the industry. This derives mainly from stringent regulatory requirements instituted after the banking crisis of the mid-1990s to enhance solvency of commercial banks. Whether this regulatory regime has benefitted Zambian commercial banks in terms of competition and efficiency performance is a matter we take up in empirical sections of this dissertation.

Chapter Three

An Overview of the Microeconomic Theories of Banking

3.1 Introduction

Until the early 1970s, the theory of the banking firm focussed on one form of bank behaviour or another without addressing the banking firm as a whole in terms of its functions, conduct, performance and response to regulatory and other policy changes (Baltensperger, 1980). In particular, the theory of portfolio selection pioneered by Markowitz (1952) and Tobin (1958) dominated the literature. Portfolio selection models emphasised the role of banks as investors without analysing their behavioural aspects as firms. This perception changed in the early 1970s with the emergence of microeconomic models of banking with roots in the industrial organisation (IO) literature. In particular, the seminal work of Klein (1971) and Monti (1972) later referred to as the Klein-Monti model provided insights into the behaviour of commercial banks under different market structures.

This part of the thesis gives an overview of some of the authoritative studies which have shaped the analysis of banking behaviour over the years. This review is not exhaustive but aims at highlighting theories relevant to the study of competition and efficiency in the banking industry in order to properly contextualise the discussion of these issues in this study. Whilst acknowledging the contribution of early theories of banking vis-à-vis portfolio selection and related models, a particular point of emphasis is placed on those models founded on microeconomic theory. Studies by Pyle (1971), Santomero (1984), Bhattacharya and Thankor (1993) give a comprehensive review of early portfolio selection models. For a review of both portfolio selection and microeconomic models of banking, see Baltensperger (1980) and Swank (1996). Freixas and Rochet (1997) is a particularly influential source of banking theories and models founded in the IO literature.

A starting point for the review is recognition of the banks' basis for existence and the role they perform (Allen & Santomero, 2001). This issue has been touched on in the introductory chapter of the thesis. However, to reiterate, the basis for banks' existence is to ameliorate risks associated with market imperfections. Specifically, because of information asymmetry and associated problems of moral hazard and adverse selection in credit markets, banks become an important avenue of overcoming such asymmetries through their intermediation role. It is this very special function which distinguishes banks from other firms and has been the central theme of theoretical as well as empirical research on banking.

3.2 Theories of banking

Theories of banking are directly linked to the banks' basic functions in an environment characterised by imperfections. Baltensperger (1980) and Swank (1996) classify these theories as falling within two broad categories. The first category is the incomplete models of banking (e.g., portfolio selection and/or risk management theories, models of bank runs, credit rationing models and bank regulation models). The second category is called complete theories of banking, principally based on the industrial organisation literature. It is important to emphasise that the multiplicity of different theories of the banking firm is attributed to the inability of any one model to address all aspects of banking. Below is a brief discussion of some of the models of banking.

3.3 Partial theories of banking behaviour

3.3.1 Portfolio models of banking

Portfolio selection bank models use risk management as the basis for analysing banking behaviour. Pioneered by Markowitz (1952) and Tobin (1958) the main thrust of portfolio selection theory is that banks select a portfolio of securities rather than individual assets (and liabilities) based on risk-return characteristics and treat these markets as perfectly competitive. According to Markowitz (1952) asset diversification defines the optimal investment decision for a financial firm, implying that banks behave as rational investors and decide upon the optimal portfolio mix to maximise expected profits.

In the Tobin (1958) framework, liquidity preference of households and firms forms the basis of the analysis. The framework includes the risk free asset and analyses the role of debt contraction in portfolio choice. Miller and Modigliani (1961) and Sharpe (1963) built on this theory, completely revolutionising the field of corporate finance. The Miller and Modigliani (1961) model demonstrates that the value of a firm is not affected by the method of financing investment (that is, equity or debt). Instead, it depends on the outcome of that investment defined by the return the firm receives. On the other hand, Sharpe (1963) explored an approach known as market (or single factor) model by reasoning that the return on each security is linearly related to a single index, such as the stock market index.

3.3.1 Models of bank runs

Bank runs are a common feature of financial crises and have played a significant role in monetary history (Kindleberger, 1978). Bank runs have a tendency to be contagious with devastating economic outcomes, including contraction in output. For the United States, studies have shown that the economic damage inflicted on the economy during the Great Depression was mainly induced by bank runs (Bernanke, 1983). The financial meltdown of 2008 stirred up renewed debate on the effect of insability in the banking system and its effect on global and national economies.

In their influential paper, Diamond and Dybvig (1983) developed a framework for analysing bank runs, liquidity and optimal insurance scheme. They argue that due to withdrawal risk, banks can be forced to liquidate their assets at a loss to meet sudden liquidity requirements by depositors. To prevent a run on the banking system, the optimal solution is government regulation in form of a deposit insurance. Allen and Gale (2000) provide further evidence and show that banks hold interregional claims (similar to interbank assets) as insurance against liquidity preference shocks. When there is no aggregate uncertainty in the market the first best allocation of risk sharing can be achieved, but this arrangement is financially fragile. Therefore, it is easier to see that contagion can easily spread throughout the financial system when a shock hits one region and is transmitted to other banks. The possibility of contagion depends strongly on the completeness of the structure of interregional claims (Allen and Gale, 2000:1). Freixas and Rochet (1997) provide a good overview of models of bank runs and the role of regulation.

3.3.2 Bank regulation models

Related to models of bank runs is another class of theories of banking, namely bank regulation models. The basis of bank regulation stems from their susceptibility to failures. Therefore, regulatory authorities institute mechanisms to curtail factors that may jeopardise the proper functioning of the financial system. Since banks hold a substantial amount of depositors' funds, a small disruption to their functioning could destroy household savings and at the same time restrict firm's access to credit.

Essentially, regulation is a response to market failures and imperfections in banking. Freixas and Santomero (2002) argue that one cannot discuss banking, its existence, its regulation or the effect of such regulation in a world without financial imperfections. Therefore, financial regulation instils discipline by compelling banks to behave in a prudent manner. Different bank regulation models have been proposed in the literature. A dominant view emphasises the safety of depositors' funds and soundness of the banking system (Santos, 2000). Theoretically, this model prescribes the minimum capital or equity requirements aimed at inducing optimal governance of banks. Hence, capital standards are used to define the threshold for the transfer of control from shareholders to the regulator. The safety and soundness regulation model also prescribes the minimum level of assets held in one sector or firm, highlighting the importance of assets diversification and reduction of funding risk.

The other model of bank regulation stipulates the entry requirements and activity lines of banking business. Increasing or decreasing the cost of entry into the banking sector could affect the profitability of firms already competing in the industry. It could also determine the degree of competition and level of efficiency in the banking sector. When bank regulation restricts the banks' scope of permitted activities under a given charter, this has important implications on bank conduct and the evolution of market structure. Entry regulatory policy may produce unintended outcomes. Therefore, it must be well designed and made compatible with competitiveness of the industry (Crampton, 2003).

3.3.3 Credit rationing models

The prevalence of information asymmetry and associated problems of moral hazard and adverse selection in credit markets has been the central area of research for many economists. The problems of moral hazard and adverse are especially acute in the banking industry and interact with other banking characteristics to determine bank performance.

Hodgman (1960) was one of the first to develop a credit rationing theory consistent with profit-maximizing behaviour. This framework has remained at the centre of the literature of credit-rationing models. The basic notion is that the bank's risk of loss (risk of default) is positively related to loan exposure. Hodgman (1960) envisages a bank's expected return as consisting of two parts with probabilities attached to them. The minimum return is realised when there is default and, in the absence of default, the full return is obtained and it is given by the loan rate less the cost of raising deposits on the money market. At low levels of loan amounts, the probability of default is negligible, but rises with an increase in the amount of loans. As the loan size increases after a certain point, the probability of default rises so that the profit on the loan starts to decrease. This results in a backward bending loan offer curve.

The widely cited model of credit rationing, due to Stiglitz and Weiss (1981), is similar in reasoning to Hodgman's (1960) characterisation. Stiglitz and Weiss (1981) show that because of moral hazard, a high loan interest rate induces borrowers who succeed in obtaining loans to engage in risky projects with a high probability of failure but promising a high return if successful. Similarly, because of adverse selection, low quality investors with a high risk of default seek to obtain loans even at a higher interest rate. Therefore, an increase in the loan rate above some critical level chases away good borrowers who are unable to repay the loan at a higher rate. To exclude less creditworthy customers, banks set the loan rate below the market clearing threshold and rations credit at this rate, driving away some loan applicants. In equilibrium, credit rationing is an optimal response to moral hazard and adverse selection. For an overview of credit rationing models see Freixas and Rochet (1997) and Santomero (1984).

Incomplete models discussed above do not provide an integrated view of banking behaviour, including an endogenous determination of the total scale of operation of the banking firm (Baltensperger, 1980). These models also exclude from the analysis the importance of real resource costs banks incur when performing their functions. For example, in screening their customers, banks expend an enormous amount of resources, related to labour, capital and other material costs (Swank, 1996; Hanak, 1992; Baltensperger, 1980). Many other bank functions also have implications on conduct and performance and the structure of the banking sector. For this reason, a partial view of the banking firm is inconsistent with the analysis of banking behaviour. Hence, Murphy (1972) admits that "*my own work dealt with most aspects of banking….but in each case a partial view of a particular problem was taken, a procedure about which I felt somewhat uneasy"* (p.614).

In view of the partial approach taken by incomplete models of banking, Swank (1996), Hanak (1992) and Baltenesperger (1980), among others, offer alternative explanations for modelling the banking firm. They argue that complete models with foundation in industrial organisation economics better describe the nature and performance of banks. Complete theories of bank behaviour are a subject of the subsequent discussions.

3.4 Complete theories of banking

The production of bank output requires use of some inputs, which entails incurring both financial and real resource costs. Intuitively, banks intermediate between borrowers and lenders and therefore incur transactions and monitoring costs. Banks also incur screening costs due to information acquisition on the quality of borrowers and their collateral. Therefore, a model of banking must incorporate the real resource costs incurred by banks in the intermediation process. These real resource costs include expenses on labour, equipment and services rendered to the bank customers, including those incurred on settlement of payments. One would therefore think of a service-based theory of banking. Complete theories of banking have been dominated by models founded in industrial organisation literature. Some of these models are discussed below.

3.4.1 Industrial organisation models of banking[17]

Models of banking that use the theory of industrial organisation as their foundation have for a long time received wide application in the banking literature. These models deal with growth, size, market structure, conduct and performance (efficiency) in the banking industry. Previous models of the banking firm ignored the extent of rivalry in banking markets and the number of banks and relative size were taken as exogenously determined (VanHoose, 1988). Therefore, because individual banks interact strategically with other banks in a manner that influences their behaviour, complete models of banking have emerged to fill the void left by incomplete models. This reasoning derives from the fact that although special in nature, banks possess behavioural characteristics similar to those of non-financial firms. Therefore, a theory of the firm suited to banking can be applied to analyse their behaviour. In view of this, IO models have formed the basis for analysing the bank production process in the context of the real resource cost theory of the banking firm.

[17] This section borrows significantly from Freixas and Rochet (1997).

3.4.2 Banking behaviour under perfect competition

The workhorse of complete models and indeed many of the IO approaches to banking is the Klein-Monti model (Klein, 1971; Monti, 1972). In a perfect competition framework, the model assumes that the banking system is composed of a large number of banks each of which is an atomistic individual unable to influence market price and/or output. Therefore, banks are profit maximising agents taking price as given while minimising costs. The model is set out as follows. There are N small homogenous banks, with an identical cost function denoted by $C = C(D,L)$ depicting the production costs of deposits (D) and loans (L), respectively. As price takers in the market, banks take the loan rate (r_L), deposit rate (r_D) and the money market rate (r) as given.

The individual bank faces a downward sloping demand for loans $L(r_L)$ and an upward sloping supply of deposits $D(r_D)$. Inverting these curves yields $r_L(L)$ and $r_D(D)$ implying that the rate of interest on loans is a function of loan demand and the deposit interest rate is a function of supply of deposits, respectively. The cash reserves, R are given by the equation below

$$R = (1-\alpha)D - L \tag{3.1}$$

where, α denotes fraction of deposits held in cash (unremunerated) reserves at the central bank.[18] The bank also participates in the interbank market and can lend or borrow at the interest rate struck by interbank liquidity conditions. The bank's profit maximisation objective is therefore given by

$$\pi(L,D) = (r_L - r)L + \left[r(1-\alpha) - r_D\right]D - C(L,D). \tag{3.2}$$

The banks' profit (π), is a sum of intermediation margins on loans and deposits, net of production costs. The first order conditions are given by Equation (3.3):

[18] In some countries, central banks pay interest on cash reserves. Furthermore, some developed countries have done away with required reserves altogether. In Zambia, cash reserves do not attract interest payment.

$$\frac{\partial \pi(L,D)}{\partial L} = (r_L - r) - C_L^{'}(L,D) = 0$$

$$\frac{\partial \pi(L,D)}{\partial D} = r(1-\alpha) - r_D - C_D^{'}(L,D) = 0 \tag{3.3}$$

where $(r_L - r)$ is the margin between the loan rate and interbank rate, $r(1-\alpha)-r_D$ is the margin between the deposit and interbank rate, $C_L^{'}$ is marginal cost of producing loans and $C_D^{'}$ is marginal cost of servicing deposits. Intuitively, Equation (3.3) states that a competitive bank will adjust its volume of loans and deposits at the point where interest rate margins equal marginal costs. This means that an increase in the deposit rate will cause a reduction in the bank's demand for deposits. Conversely, a high loan rate leads to a fall in the supply of bank loans. Rearranging Equation (3.3) and solving for r_L yields the following result:

$$r_L = \alpha r + r_D + C_L^{'}(L,D) + C_D^{'}(L,D) \quad (3.4).$$

Upon further rearranging we obtain the interest spread given by Equation (3.5) below

$$r_L - r_D = \alpha r + C_L^{'}(L,D) + C_D^{'}(L,D) \quad (3.5).$$

The result given by Equation (3.5) is a fundamental finding. It states that the intermediation margin $(r_L - r_D)$ is driven by the reserve ratio and the sum of marginal costs of servicing loans and deposits by the bank. The central message implied by the perfect competition model of banking is that the intensity of competition among banks should induce a reduction in the price of bank products and services while quantity should increase. Consequently, this would dampen monopoly profits enjoyed by individual banks (Besanko & Thakor, 1992; Guzman, 2000).

One of the applications of the perfect competition model of banking is a study by Alencar and Nakane (2004). Using Brazil as an example, the authors analyse banking behaviour in a model with agency costs and costly verification technology. The model evaluates two cases – a competitive scenario and a case where banks have market power. Numerical simulation results show that greater competition in the loan market enhances the response of the real economy to an interest rate shock. The authors also submit that higher competition and/or a more efficient verification technology reduces the reaction of both the default rate and interest rate spread to an interest rate shock. Finally, they argue that the influence of the verification technology in the economy's dynamic response is greater when banks operate under perfect competition. This result gives credence to the hypothesis that monetary transmission is strong under highly competitive banking market conditions.

3.4.3 Klein-Monti original monopoly model of banking

The theory of perfect competition in banking is one polar of market structure in the industrial organisation literature. Given the assumptions of price taking behaviour, this model is very restrictive and unrealistic for practical purposes. Furthermore, because the banking sector is characterised by different forms of barriers to entry, the imperfect competition model would probably be the more appropriate model of analysis. Before presenting the oligopoly model of banking, we need to discuss the original formulation of another extreme case of banking, namely the monopoly model of bank behaviour.

The framework is simple in structure yet it has served as benchmark for analysis of bank behaviour. The model assumes a single bank operating in the industry, choosing among three assets: cash reserves, loans and Treasury bills. The model also distinguishes among three types of liabilities, namely demand deposits, time deposits and equity capital which is set exogenously. This model establishes that a profit maximizing monopoly bank operating in a well-developed financial market will equalise the marginal cost of deposits and the marginal return on loans to an exogenous market rate such as the Treasury bill or the interbank rate. Reasoning analogously as in the perfect competition framework the monopolist bank maximises[19]

$$\pi(L,D) = r_L(L)L - r\left(D(1-\alpha) - L\right) - r_D(D)D - C(L,D) \tag{3.6}$$

Assuming that the profit function is concave and twice differentiable, the following first order conditions obtain

---

[19] The difference between the perfect competition and monopolist banking case is that the latter takes into account the infuence of the stock of loans on the loan rate $r_L(L)$ and similarly for deposits, $r_D(D)$.

$$\frac{\partial\pi(L,D)}{\partial L}=r_L^{'}(L)+r_L-r\text{ - }C_L^{'}(L,D)=0 \tag{3.7a}$$

$$\frac{\partial\pi(L,D)}{\partial D}=-r_D^{'}(D)D+r(1\text{ - }\alpha)\text{ - }r_D-C_D^{'}(L,D)=0 \tag{3.7b}$$

From the above equations we have the prediction of the Klein-Monti monopolistic model of banking which states that the optimal amount of loans (L^*) is determined solely by Equation (3.7a) and the optimal stock of deposits (D^*) is determined by Equation (3.7b). Similarly, the optimal solutions for loan and deposit rates are given by r_L^* and r_D^*, respectively. Therefore, the bank's lending decision $L(r_L)$ is independent of the deposits supply function $D(r_D)$.

Let the loan demand elasticity be defined as $e_L=-\frac{r_L L^{'}(r_L)}{L(r_L)}>0$ and the deposit supply elasticity as $e_D=\frac{r_D D^{'}(r_D)}{D(r_D)}>0$. Substituting these terms in (3.7a) and (3.7b) and using the equilibrium solutions of r_L^* and r_D^*, we get:

$$\frac{r_L^*-r-C_L^{'}}{r_L^*}=\frac{1}{e_L(r_L^*)} \tag{3.8a}$$

$$\frac{r(1-\alpha)-r_D^*-C_D^{'}}{r_D^*}=\frac{1}{e_D(r_D^*)} \tag{3.8b}$$

The left hand side of expressions (3.8a) and (3.8b) depict the Lerner Indices for loans and deposits, respectively and state that a monopolistic bank sets its volume of loans and deposits such that the Lerner Indices equal inverse elasticities.[20] These Lerner Indices are adapted to the banking sector to

[20] The Lerner Index (LI) is given by $(P\text{ - }MC)/P$ where P is price of bank output and MC is marginal cost of production.

depict market power exercised in the loans and deposits markets. The greater the banks' market power in these markets, the smaller the respective elasticity and the higher the Lerner Index.

Although the monopolist model of banking yields some useful properties, it has received a lot of criticisms, centred mainly on the independent dichotomy of loans and deposits determination. For example, Swank (1996) argues that this dichotomy is a direct consequence of simplifications inherent in the model, rather than a result of price setting behaviour. Not surprisingly therefore, extensions to the model suggest that assets and liabilities and interest rates are all interdependent.

For instance, Dermine (1984; 1986) and Goodman and Santomero (1986) examined the implications of different premium structures of deposit insurance on the model. The results indicate that incorporating liquidity and solvency factors invalidate the assumption of the independence dichotomy. The assumption of risk neutrality has also been questioned. Given the uncertainty in financial markets, risk is a fundamental problem in banking. Therefore, an explicit treatment of risk in the model exposes the key role of market imperfections and yields results ignored by the Klein-Monti model (Pringle, 1974).

3.4.4 The imperfect (oligopolistic) competition model of banking

The disillusionment with failures of perfect competition and monopolistic models to provide practical explanations of banking behaviour has naturally yielded to different models of banking. For all practical intents, imperfect competition more reasonably approximates actual banking behaviour. There is a voluminous literature in support of this view.[21] As conjectured by proponents of the theory of oligopolistic markets, banks typically operate in a monopolistically competitive environment, making this class of imperfect competition models more appealing. Therefore, models of imperfect competition in the banking sector offer robust improvements to the analysis of bank behaviour (Freixas & Rochet, 1997; Toolsema, 2003).

A more accurate description of reality is to assume that the banking sector is characterised by a finite number of banks, N operating in an imperfect Cournot competition framework. For simplicity,

[21] In Chapters four and six below, we give a detailed review of the literature for competition and market power in the banking sector and outline the main conclusions contained in the studies reviewed. Empirical evidence is then provided based on the Zambian banking sector.

the basic assumptions of the traditional Klein-Monti model remain unchanged. However, all banks now face the same linear cost function denoted by

$$C_n\left(L,D\right)=\gamma_L L+\gamma_D D \quad n=1,......,N \tag{3.9}$$

Furthermore, each bank takes the amount of loans and deposits chosen by other banks as given and maximises its profit by choosing its own amount of loans and deposits. Therefore, the profit maximisation condition for each bank is

$$\pi=\left\{\left(r_L\left(L_n+\sum_{m\neq n}L_m\right)-r\right)L_n+\left(r\left(1-\alpha\right)-r_D\left(D_n+\sum_{m\neq n}D_m\right)D_n\right)-C\left(L_n,D_n\right)\right\} \tag{3.10}$$

Solving the objective function, Freixas and Rochet (1997) show that a Cournot equilibrium is an $N-$ firm of vectors of the stock of loans and deposits $\left(L_n^*,D_n^*\right)_{n=1,2,..,N}$, so that for every n, $\left(L_n^*,D_n^*\right)$ maximises the profit for each individual bank. The solutions yield a unique equilibrium where each bank sets its equilibrium stock of loans, $L_n^*=\dfrac{L}{n}$ and volume of deposits, $D_n^*=\dfrac{D}{n}$. The first order conditions with respect to the equilibrium stock of loans, L_n^* and optimal level of deposits, D_n^* are given by the following set of equations:

$$\frac{\partial\pi_n}{\partial L_n}=r_L^{'}\left(L^*\right)\frac{L^*}{N}+r_L\left(L^*\right)-r-\gamma_L=0 \tag{3.11a}$$

$$\frac{\partial\pi_n}{\partial D_n}=-r_D^{'}\left(D^*\right)\frac{D^*}{N}+r\left(1-\alpha\right)-r_D\left(D^*\right)-\gamma_D=0 \tag{3.11b}$$

Rewriting the first order conditions in elasticity form yields the equivalent Lerner Indices as follows:

$$\frac{r_L^*-\left(r+\gamma_L\right)}{r_L^*}=\frac{1}{Ne_L\left(r_L^*\right)} \tag{3.12a}$$

$$\frac{r(1-\alpha)-r_D^*-\gamma_D}{r_D^*}=\frac{1}{Ne_D\left(r_D^*\right)} \tag{3.12b}$$

The equilibrium solutions of the oligopoly model differ from those obtained in a monopolistic banking model only in the measurement of elasticities. For the oligopoly framework presented above, the elasticities are multiplied by a factor of N, the total number of banks in the industry. Intuitively therefore, the Klein-Monti model of monopolistic banking may be reinterpreted as a special case of imperfect competition when $N=1$. When $N \to +\infty$, we have the other end of perfect competition. Intermediate N values of banks denote monopolistic competition. Accordingly, the Lerner Index (LI) closer to zero denotes intensifying competition while the converse obtains when competition among banks is weak. If a firm leaves the industry, output should fall and price rise, so that the price-cost margin (Lerner Index) increases. Thus, for $N=1$, the Lerner Index tends to approach unity because price is always above marginal cost. Under perfect competition $LI=0$ since price is equal to marginal cost. Monopolistic competition can therefore be being conceived as a middle range of the Lerner Index $(0<LI<1)$. The intermediate case of the Lerner Index measures the average deviation of banks' behaviour from the bipolar cases of monopoly and perfect competition. According to Coccorese (2009), it expresses the true degree of market power exercised by banks.

These conjectures can then be used to evaluate the responsiveness of the changes in the money market rate on the loan and deposit rates as the number of firms is altered in the banking industry. For instance, when the intensity of competition increases as the market tends to a perfect competitive case, a change in the money market rate induces a smaller response in the loan rate while the response of the deposit rate is greater under similar conditions. This can be seen by partially differentiating the solutions of the equilibrium loan rate r_L^* and deposit rate, r_D^*.The equilibrium loan rate is given by

$$r_L^*=\frac{Ne_L\left(r_L^*\right)}{Ne_L\left(r_L^*\right)-1} \tag{3.13}$$

The equilibrium deposit rate is denoted by

$$r_D^* = \frac{r(1-\alpha)Ne_D(r_D^*)}{1+Ne_D(r_D^*)} \tag{3.14}$$

The response of the r_L^* to changes in r is obtained by the following equation

$$\frac{\partial r_L^*}{\partial r} = \frac{1}{1-\frac{1}{Ne(r_L^*)}} \tag{3.15}$$

From Equation (3.15), the response of the loan rate to changes in the money market rate is determined by the intensity of competition. Hence when $N \to +\infty$, r_L^* becomes less sensitive to changes in r. The response of the deposit rate can be derived analogously (see Freixas and Rochet, 1997). Another prediction of the imperfect (oligopoly) competition model presented above is that the interest spread between the loan rate and the deposit rate narrows down as N increases due to rising competitive pressures. The observed pattern in the Zambian banking industry is at variance with the theoretical prediction, mainly due to structural and macroeconomic factors, including embellished inflation expectations and high default rates.

Although the Cournot model discussed here addresses the shortcomings inherent in the standard Klein-Monti monopoly model, there have been some extensions of this framework by introducing aspects ignored in the formulation of the imperfect competition model. For example, Prisman, et al. (1986) developed a stochastic model consistent with oligopolistic competition by introducing uncertainty and liquidity requirements using a two-stage bank decision framework. They assume that an individual bank is a price taker in the bond market and price setter in the loan market and certain deposit markets.

The model yields very interesting results. Both liquidity constraints and uncertainty push the bank to borrow from the central bank. However, since the central bank charges a penalty rate in order to encourage interbank activity, the equilibrium is such that the optimal loan and deposit rates are a function of the refinancing penalty rate at the central bank. More importantly, the model shows that under conditions of uncertainty there is interdependence between asset/liability management such that the elasticity of deposit supply lowers the loan rate by making the banks more aggressive in the loan market.

3.4.5 Real resource cost models of banking

Real resource cost models are in the spirit of Swank (1996), Hanak (1992), Santomero (1984), Baltensperger (1980) and Sealey and Lindley (1977). The models emphasise a bank's production technology taking into account all production costs (financial and non-financial) to reflect the constraints banks face in performing their primary function of intermediation. Swank (1996) submits that a complete neglect of real resource costs in a theory of banking is inadvisable. This is because operating costs account for a sizable amount of total banking costs with obvious implications on the level of production.

The intuition is that production of earning assets (loans for example) and deposit mobilisation attract costs that may constrain the bank from performing these functions. Such costs may exhibit a U-shaped average cost curve and rising marginal costs over a range of output. However, as Pesek (1970) has observed, previous models on the determination of the equilibrium quantity of money paid less attention to the role played by real resource costs in theoretical analysis of banking. Therefore, a model of the banking firm must adequately reflect the price and cost elements in the analysis of bank behaviour.

These theories are based on a specific production technology for a financial firm embodying two stages in the bank's optimisation objective. The first stage is cost minimisation, which reflects the special features of the production technology itself. In the first stage, analysts invoke duality theory and derive factor input demands using the Shepherd's Lemma. The second stage of profit maximisation results from the bank's choice of a profit maximising level of output, prices and distribution services, subject to the cost minimisation level of inputs, the demand functions for its output and the balance sheet constraint.

Real resource cost theories of the banking firm have formed the bedrock for a large body of modern empirical work in banking competition, market power and efficiency. Therefore, the different variants of the Klein-Monti models discussed above are examples of real resource cost theories of banking.

The application of real resource cost and other related imperfect competition models has largely been confined to industrial countries while their use in bank studies of nonindustrial countries has been constrained by lack of detailed bank level data. The availability of highly disaggregated data in recent years has made it imperative to test the applicability of these banking models in developing countries. In particular, the focus of the analysis has been in relating variations in bank revenues and costs to constraints imposed by structural and institutional factors in modelling bank conduct. The present study is one of the new efforts aimed at validating the evidence from developed countries by subjecting these theoretical models to banking systems of less developed economies characterised by a host of institutional bottlenecks and macroeconomic impediments. In this regard, the study opens up new research opportunities in this genre of models.

3.5 Definition of inputs and outputs for a banking firm

The issue of what constitutes inputs and outputs for a banking firm is highly controversial in the literature. Two approaches have dominated the literature on the measurement of inputs and outputs. The first, called the production approach, stipulates that banks and indeed all financial firms are not different from non-financial firms and therefore, both inputs and outputs must be measured in physical units. Thus, banks use labour, capital and other material resources (inputs) to produce the number of demand deposit accounts and quantity of loans transactions processed as outputs. In the production approach, only operating expenses arising from use of personnel, computers, etc., are taken into account and nothing else matters. Sherman and Gold (1985) motivate most of the research in the production approach. However, Sealey and Lindley (1977) and Berger, et al., (1994) argue that the use of production approach stems from a lack of proper distinction between technical and economic production and this difference is critical for commercial banks and all financial intermediaries. To this end, an alternative view to the measurement of bank inputs and outputs has been advanced. This view is based on the premise that banks engage in the transformation process involving the borrowing of funds from surplus spending units and lending those funds to deficit spending units. This process is known as financial intermediation. This special function is an important distinguishing feature of banks from non-financial firms (Gurley & Shaw, 1955). The transformation process creates a product more highly valued than the original input elements. Therefore, the concept of production in the economic sense takes into account the banks' objective of maximising profits by creating value to output measured at market prices.

For commercial banks, services made to creditors are more associated with the acquisition of economic inputs (labour, capital, materials and loanable funds) since these services require banks to incur real resource costs without yielding any direct revenue. On the other hand, as intermediaries, banks use labour, capital, materials and borrowed funds to produce earning assets which generate revenue for the bank (Sealey & Lindley, 1977). Accordingly, banks incur costs related to usage of labour, capital and materials involved in producing different services to depositors. They also make interest payments to depositors for their funds. These expenses constitute the banks' input costs.

The intermediation approach has more appeal to a developing country such as Zambia where banks continue to perform the basic function of financial intermediation. Accordingly, and in line with previous research, this study uses the intermediation approach on the following strengths. Firstly, bank deposits and other borrowed funds constitute the main sources of funds for Zambian commercial banks, accounting for more than 70.0 percent of total liabilities. Therefore, deposits accurately reflect banks' inputs rather than outputs, especially that there is a limited range of services banks pay to deposit holders. Secondly, the intermediation approach incorporates all expenses (interest expenses and non-interest expenses) and is therefore consistent with real resource cost models. The added advantage of the intermediation approach is that data on monetary values of assets and deposits are more readily available than the deposit and loan accounts, the key data required for the production approach.

Chapter Four

Cost Efficiency and Scale Economies in Zambian Banking

4.1 Introduction

Research on banking efficiency in less developed countries (LDCs) has received little attention despite rapid growth in this literature in other parts of the world. This is rather unfortunate given the dominance of the banking sector in the financial system of many of these LDCs. Bank efficiency performance is a prerequisite for economic growth and this impact is strong and clear-cut, since inefficiency is tantamount to wastefulness which leads to lower profits or higher bank tariffs or both (Bikker & Bos, 2005). From the perspective of a practising banker, the interest in bank efficiency is made more relevant by the changes in the banking industry, spurred by technological advances and other economic developments. Therefore, banks and regulators should use efficiency evaluation and benchmarking as methods to identify the best practices in order to improve the performance and productivity of commercial banks.

Banks play an invaluable role in ameliorating the effects of information asymmetry by transforming risk, size and maturity of financial instruments. In doing so, they convert short-term liabilities into long-term productive assets. Indeed, economies with a sophisticated and efficiently performing financial system are more capable of evaluating assets whose real value would otherwise be difficult to determine in the absence of properly functioning financial institutions (Thakor, 1996). For this and other reasons, regulatory authorities and bank managers alike must aspire to achieve operating efficiency which is essential for vibrant economic performance and improvement in social welfare. Cost efficiency is also beneficial to consumers because it translates into lower prices and provision of high quality bank services.

In light of the above, a number of transition and developing economies embarked upon financial liberalisation policies, seen as a panacea for stimulating efficiency performance after many years of interventionist policies. Without doubt, the reform measures injected fresh air in the functioning of the banking sector. In Zambia specifically, institutional and regulatory reforms significantly shaped the financial landscape and impacted greatly on banks' profitability, cost structure and efficiency. However, the reforms also posed great challenges for general bank conduct resulting in little progress being made in the area of efficiency. High cash reserve ratios and other regulatory requirements have also been blamed for the low level of efficiency. These weaknesses have been compounded by the presence of widespread distortions elsewhere in the economy such as poor quality of collateral and balance sheet weaknesses for bank borrowers, which have been reflected in high proportion of nonperforming loans to total loans, estimated at 13.6 percent between 1998 and 2006. To circumvent credit losses arising form toxic assets, banks have responded by holding large amounts of government securities and accumulating off-shore deposits in foreign currency. This has reduced the amount of credit available to the domestic economy, exacerbating financial disintermediation.

In Zambia, the authorities have long bemoaned the low level of bank efficiency as an impediment to financial development. In 2002, the IMF and the World Bank conducted an assessment of the financial system and observed that Zambian banks were cost inefficient as shown by wide interest rate spreads. However, as alluded to earlier, interest margins are only indicative and do not provide a comprehensive assessment of bank efficiency performance. Thus, to our knowledge, there has been no empirical study conducted to investigate the pattern of bank cost efficiency in order to justify the concerns raised above. Against the background of unsatisfactory performance of the Zambian banks on efficiency grounds, the task in the present chapter is to assess empirically, the level of bank cost efficiency and its determinants. The study is both timely and relevant in light of the recommendations articulated in the FSAP and the more recent FinMark Trust studies.

4.2 Research questions and hypotheses

The banking crisis of the 1990s changed the way the Bank of Zambia conducted its regulatory policy. After the failure of many banks, the central bank improved its supervisory capacity and implemented a new banking legislation to prevent future crises. In this streak, bank restructuring took centre stage and was aimed at strengthening banks' capital power. Capital and other regulatory requirements are important in shaping banks' behaviour because of imperfections characteristic of financial markets. Regulations take different forms, but the important ones are capital and cash reserve requirements which are imposed to protect banks from insolvency. The authorities may also impose other restrictions on banking activities, such as limits on branching and foreign exchange trading. Thus, it would be postulated that due to the presence of market imperfections, regulatory requirements attenuate the information problem by limiting risk taking in the banking industry.

However, in complying with these regulatory requirements, banks incur additional costs which may affect their performance. It turns out that these forms of regulatory requirements may affect the level of efficiency because they act as fixed inputs in the production process. To illustrate, tighter capital regulatory requirements meant to safeguard the banking industry may produce some unintended adverse consequences on bank behaviour, including inefficient allocation and use of resources. Therefore, in determining bank cost behaviour, it is important to incorporate regulatory factors (Elliehausen, 1998).[22] Research by Fare, et al. (2004) found that risk-based capital standards had a significant impact on profit efficiency. Hughes and Mester (1993) also argue that an increase in reserve requirements raises the opportunity cost of funds and acts as a tax on the price of deposits, thereby impairing banks' decision making. Estrada and Osorio (2004) offer one option to correct for the effect of regulatory intensity by explicitly allowing financial capital costs to determine the level of efficiency. This approach is aimed at directly capturing regulatory burden borne by the banks and helps understand the dynamics of hidden costs which may be difficult to observe in practice.

Another important consideration relates to the macroeconomic environment under which banks operate. Significant changes in the business cycle and general economic performance and the policy environment all have implications for bank efficiency performance. For instance, in Zambia, persistent government fiscal deficits, high inflation and rapid currency depreciation created uncertainty in the economy. Collectively, these factors may have affected resource allocation in the banking sector.

In view of the foregoing, this chapter seeks to answer the following research questions. What has been the pattern of cost efficiency of Zambian banks since the end of the crisis? Do loan quality and other bank-specific factors intensify or attenuate bank inefficiency? Furthermore, is regulatory burden a major source of cost inefficiency in the Zambian banking industry? Are cost efficiency and bank ownership structure related, and if so, which bank categories are more efficient than others? These are important policy questions which would shed light on the nature and dynamics of efficiency performance of Zambian banks since the banking crisis of the 1990s.

---

[22] Some studies have estimated direct and compliance costs of regulation see for instance, Franks, *et al.* (1997) and Elliehausen (1998) for a survey. These estimates show that costs of regulation range between 9 % – 13 % of non-interest expenses. Studies also show that there are economies of scale associated with regulatory costs; large banks incur lower per unit regulatory costs than smaller banks. Our analysis reveals that the income statement from Zambian banks does not report costs associated with regulatory requirements. Therefore, it is very difficult to estimate the extent of regulatory costs and apply them to banking efficiency. To circumvent this handicap in our empirical framework below, we use reasonable proxies to capture regulatory intensity.

In the remainder of this chapter, we outline in detail the approach used to answer these questions. Section 4.3 provides a theoretical foundation of economic efficiency and its application to the banking sector. In Section 4.4 a review of relevant literature on bank efficiency is presented and discussed. The analytical framework and methodological approach taken to analyse cost efficiency of Zambian commercial banks is a subject of Section 4.5. Empirical results are presented and discussed in Section 4.6 while Section 4.7 provides a summary of research findings on cost efficiency performance.

4.3 Theoretical foundation of economic efficiency

The concept of economic (productive) efficiency is rooted in neoclassical microeconomic theory, which focuses on resource allocation and utilisation. It advocates for non-wastage of resources by emphasising cost reduction while producing the maximum possible level of output for a given technology and available inputs. The main driving force behind economic efficiency is value creation. In the process of transforming inputs into some output value, a change that increases value is an efficient change and one that decreases value is an inefficient change. For purposes of policy intervention efficiency has often been used to evaluate the effectiveness of policy alternatives.

Economic efficiency is better explained by profit maximisation (or analogously, cost minimisation) but is often associated with perfectly competitive markets than with monopoly because the latter leads to deadweight loss. For firms operating in a competitive industry, efficiency gains accrue when firms earn only normal profits in the long-run and respond to changes in consumer preferences by increasing output. Whether this output in sold at the same, higher or lower price depends in large measure on the position of the cost curves in the long-run.

Economic efficiency also encompasses allocative efficiency, which occurs when a firm's inputs are allocated in such a way as to maximise its benefits (profits, revenue and output) depending on the firm's objective function. In a perfectly competitive market, allocative efficient outcomes occur when price is equal to marginal cost. Allocative efficiency also addresses the issue of the right mix of inputs and quality of output produced.

Finally, X-efficiency, introduced by Leibenstein (1966) refers to efficiency in production by linking inputs to outputs. It is an economic expression for the effectiveness with which an organisation uses its given set of inputs to produce outputs. Specifically, it refers to the internal organisation of firms and its response to external factors. Under such circumstances, both motivational factors (e.g., moral and bureaucratic inertia and human errors) and competitive pressures may affect X-efficiency. In many of his writings, Leibenstein repeatedly argued that X-efficiency was superior to allocative efficiency, implying that the latter effect was trivial.

Evidently, the concept of efficiency is broad and has been assigned different interpretations. The concept of X-efficiency is especially controversial. Since Leibenstein first proposed the term, analysts have sought to evaluate its meaning and in the process, different interpretations have emerged. For instance, Peel (1974) contends that cost reduction can be observed in a movement from monopoly to perfect competition, not due to X-efficiency but because managers are forced to give up goals inconsistent with cost minimisation as the environment becomes more competitive. Equally, Stigler (1976) cautions that failure to recognise higher than minimum cost levels are mainly due to rationally calculated utility maximising strategies (including leisure and expense preference) by workers rather than by X-inefficiency.

The above efficiency definitions refer to static efficiency. The alternative definition derives from dynamic efficiency (Jameson, 1972). Dynamic efficiency takes into account change in technology as a main source of productivity, so that over time, the production frontier shifts due to technological advancements. Therefore, according to Abel, et al., (1989), dynamic efficiency could be associated with long-run economic growth arising from productivity change due to capital accumulation.

Scale economies and economies of scope provide another perspective for analysing firm efficiency performance. Economies of scale occur when more units of a good or service can be produced on a larger scale, yet with less input costs. Therefore, economies of scale are associated with size of the firm, implying that larger firms enjoy economies of scale due to larger production technology. For a long time, economic growth has been explained by the theory of economies of scale. Marshall (1961) distinguished between internal and external economies of scale, attributing the former to lower costs which lead to higher production and the latter as emanating from external forces such as improvements in the transportation network. External economies of scale may thus benefit all firms in the industry by lowering costs and stimulating firm growth.

In contrast, economies of scope relate to a reduction in costs resulting from joint production. Thus, a firm which decreases its average costs because of changes in production of different products (related or unrelated) is said to be enjoying economies of scope. Economies of scope also provide firms with means to generate operational efficiencies, especially when these are driven by diversification. In the case of banking, for instance, it may be economical for a bank to diversify into different areas of financial services such as investment banking, commercial banking, leasing and factoring, and life insurance rather than pursue traditional commercial banking alone. Such a strategy may be driven by synergies of knowledge of commercial banking and investment banking of particular corporate customers. For larger banks, diversification may be an important business strategy aimed at reducing portfolio risk and hence gain efficiency.

Although scale economies may be important, bigger may not necessarily be better in the sense that when firms expand in size, the chain of command also becomes cumbersome. Therefore decisions made by top managers may take longer and information distorted by the time it reaches the bottom ladder of the firm. When this happens, the firm may not be realising cost reductions as predicted by theory. Accordingly, diseconomies of scale may set in. Thus, smaller firms may be more efficient than larger firms in this regard.

A similar interpretation may be offered for economies of scope. While economies of scope often provide an incentive to expand product lines, the creation of new products is often less efficient than expected, resulting in diseconomies of scope. This is because the introduction of new product brands may entail additional managerial expertise or personnel, higher raw material costs, a reduction in competitive focus, and the need for additional facilities, which collectively could result in high per-unit costs. Therefore, economies of scope may not be obvious. For the banking example above, managers may find it challenging to manage insurance business while also trying to provide commercial banking services to a wide range of customers. For this reason, the envisaged synergies may actually lead to huge cost increases.

Different factors may explain cost inefficiency in a firm. Some of the factors may be inherent in the internal organisational structure of the firm and therefore controllable. These factors may be a function of administrative mismanagement and lack of managerial expertise, inexperience of workers and poor skills levels, non-optimal diversification of assets, misallocation of inputs, among others. These factors may impair the firm's capacity to optimally respond to changing incentives and constraints. External factors, also called environment factors, may also affect the banks' inefficiency. These include regulatory constraints, macroeconomic shocks, real business cycles, strikes and labour disputes and structure of the market in which the firm is operating. Taken together, these factors may account for substantial variations in firms' performance levels. Internal factors are firm specific and therefore within the control of the firm, environmental factors are outside the control of the firm but could have a profound impact on the bank's inability to operate optimally. Therefore, in assessing the efficiency of firms, care must be exercised in differentiating between internal and external factors.

4.4 Review of the literature

The concept of economic efficiency was first mooted in Farrell's (1957) seminal work in which he showed that overall efficiency can be decomposed into "price (allocative) efficiency" and "technical efficiency". Farrell's (1957) definition of technical efficiency led to the evolution of different frontier methods for estimating efficiencies of firms. The Farrell framework gives different ways in which a productive unit can be inefficient. A decision making unit can be inefficient either by obtaining less than the maximum output available from a pre-determined level of inputs (technically inefficient) or by not using the right input mix given their prices and marginal productivities (allocatively inefficient). Although frontier approaches possess some similarities, there are significant methodological differences among them. Therefore, the approaches used for estimating economic efficiency can be categorised broadly as being either parametric (stochastic) or non-parametric (linear programming), depending on the specification and estimation of the efficiency frontier and assumptions made about the distribution of the error component.

The stochastic frontier approach (SFA) was simultaneously but independently developed by Aigner, et al. (1977), Meesen and van den Broeck (1977) and Battese and Corra (1977) to estimate production efficiency for cross-sectional data. Subsequently, Ferrier and Lovell (1990) applied the methodology to banks. The SFA involves parameterising the econometric relationship between the level of inputs and the technically efficient level of output by choosing a particular functional form for the production/cost function. The hypothesised functional form allows for a composite error term which incorporates the efficiency term as well as the effects of exogenous shocks (random error) beyond the control of the analysed units. The SFA also takes into account any possible measurement errors.

The main criticism of the stochastic frontier approaches stems from the apriori specification of the function form. Critics argue that when the specification of the efficiency function and the stochastic term are assumed a priori, any misspecification of the estimated econometric model could contaminate the efficiency scores.

Another variant of frontier estimation techniques is founded in the so-called non-parametric approaches based on data envelopment analysis (DEA) or linear programming techniques following the seminal work of Charnes, et al. (1978). The approach by Charnes, et al. (1978) uses Farrell's (1957) concept of efficiency under constant returns to scale (CRS). Later reformulation of the DEA model by Banker, et al. (1984) showed that overall efficiency can be divided into 'pure technical' and 'scale' efficiency and suggested that firms may in fact be characterised by variable returns to scale (VRS). The DEA also decomposes overall efficiency into technical and allocative efficiency.

Data envelopment analysis does not explicitly make any assumptions regarding the functional form of the frontier but empirically builds a best-practice function from observed inputs and outputs (Favero & Papi, 1995). However, a major criticism levelled against the DEA methodology is that it assumes absence of measurement error and statistical noise. Accordingly, errors are taken as measures of inefficiency. As Herrero & Pascoe (2002) have observed, the estimated efficiency scores from a DEA model may be biased if the production process is largely characterised by stochastic elements. In contrast, since the stochastic frontier model takes into account random shocks outside the decision maker's control in the measurement of efficiency, it presumably represents an improvement over the DEA. Nonetheless, both approaches have received wide application in the banking system.

4.4.1 Empirical Literature

This section reviews and discusses some of the related empirical literature on measurement of bank efficiency. The estimation of bank efficiency has developed along the lines of the two distinct frontier models-DEA and SFA and variants thereof. Traditionally, technical efficiency in banking was measured using the production function by estimating scale and scope efficiencies. However, according to Kwan and Eisenbeis (1996) technical efficiency is only a component of overall economic efficiency. Therefore, scale and scope efficiency measures are of little economic significance for financial institutions. Berger, et al. (1993) and Berger and Humphrey (1991) have argued that in fact scale and scope inefficiencies are less important than X-inefficiency in the banking industry. Similarly, Bauer, et al. (1998) argue that for policy purposes, economic efficiency is a much broader concept than technical efficiency in the sense that the former encompasses the latter and involves an optimal choice of inputs and/or outputs based on the reactions to market prices.

The empirical literature on the measurement of bank efficiency has been dominated by research from industrial countries, particularly the United States and Europe (Berger & Humphrey, 1991; Aly, Grabowski, Pasurka, & Rangan, 1990; Rangan, Grabowski, Aly, & Pasurka, 1988; Sherman & Gold, 1985). In recent years, there have been a growing number of studies for emerging markets and countries in transition. For developing countries, the literature is sparse and largely concentrated in East and South Asian countries. Berger and Humphrey (1997) provide a survey of some of the early works in developed countries for bank efficiency while Berger (2007) provides the analysis of efficiency estimates from an international perspective.

A particularly influential study of efficiency in banking in industrial countries is Berger and Humphrey's (1991) study. Berger and Humphrey (1991) argued that X-inefficiencies were in the order of 25 percent or more of overall cost inefficiencies and were more important than scale and scope economies, which accounted for only 5 percent or less. An earlier study by Rangan, et al. (1988) had found that US banks were about 30 percent inefficient largely due to pure technical inefficiency. Estimates for scale inefficiencies were shown to be very small with almost all banks operating under a constant returns to scale technology. However, Altunbas, et al. (2000) showed that in Japan, scale inefficiencies dominated X-inefficiencies. More recently, Valverde, et al. (2007) show that correcting for banks' external business environment and productivity indicators at industry level significantly reduces cost inefficiencies. Therefore, based on this new research, inefficiencies from scale operations and X-inefficiencies appear to be of equal weight. Evidently, one would think of industry productivity growth as emanating from firm level scale expansion and therefore, this would translate into lower than expected inefficiencies for individual banks.

The liberalisation of financial markets has had a profound impact on the bank efficiency literature in industrial countries as well as LDCs. For studies in industrial countries there is some evidence suggesting that deregulation fosters bank efficiency. For example, Sturma and Williams (2004) found that bank efficiency increased in post-deregulation period in Australia mainly due to competition resulting from diversity in bank types. Foreign banks in Australia were also more efficient than domestic banks but this did not translate into superior profits. Carbo´, et al. (2003) argue that improved post-deregulation economic conditions accounted for most of the improvement in costs and profits of Spanish banks. Their conclusion is that deregulation of financial markets is better than mergers. This evidence contrasts findings by Kaparakis, et al. (1994) who argued that the removal of restrictions on branching activities and the concomitant consolidation due to deregulation of US banking sector caused banks to operate in an inefficient manner during the transition phase. Other scholars have also found that globalisation of financial markets manifested in increased foreign bank penetration has tended to dwarf banking efficiency although this result must be interpreted in conjunction with the quality of institutions in host countries (Lensink, Meesters, & Naaborg, 2008).

Given the accelerated pace of economic reforms and globalisation of financial markets since the late 1980s, there has been a growing body of research assessing efficiency in banking and productivity performance in developing countries and emerging markets. This interest has been necessitated by availability of data in these countries and the need to evaluate the potency of financial reforms.

Emerging markets and former communist countries have provided a new ‘laboratory’ for testing bank efficiency. For the majority of studies in emerging markets, the evidence is consistent with prior expectations that liberalisation stimulates efficiency, particularly of private banks. This gives further impetus for continued reforms and restructuring of public banks (Christopoulos & Tsionas, 2001; Zajc, 2006; Fries & Taci, 2005; Rao, 2005; Staikouras, Mamatzakis, & Koutsomanoli-Filippaki, 2008; Weill, 2003; Yildirim & Philippatos, 2007). Similar findings have been reported for the Turkish banking sector. For example, Zaim (1995) and Isşık & Akçaoğlu (2006) showed that in Turkey, financial liberalisation fostered efficiency of commercial banks with a large number of them operating at an optimal scale. This indicated that financial reforms created incentives for banks to unlock their potential by implementing measures aimed at reducing costs in order to enhance efficiency performance. In contrast, Denizer, et al. (2007) and Yildirim (2002) observed that there was a slow recovery of efficiency by the Turkish banks mainly due to the unstable macroeconomic and financial environment which accompanied the reforms. This tended to complicate the cost performance of financial institutions.

Another notable focus of research on bank efficiency has been received from China and other Asian countries. Chen, et al. (2005) applied the DEA to analyse the efficiency of Chinese banks for the pre-and post-liberalisation period. They concluded that the level of efficiency increased markedly after the financial deregulation of 1995 with large state-owned banks and smaller banks benefitting more from the reforms than the medium sized banks. Further results by Fu and Heffernan (2007) based on the SFA show that Chinese banks gained efficiency ground in the first phase of the banking reforms, suggesting that deepening of liberalisation policies strengthen banks’ cost X-efficiency. A possible explanation for this may be that due to relaxation of entry barriers and other banking activity restrictions, competition intensified, causing many banks to adopt measures to improve productivity and efficiency. Lin (2005), Lim and Randhawa (2005) and Leightner and Lovell (1998) all make similar conclusions that productivity and efficiency improved after the reforms with the private banks the major beneficiaries from the reform process. However, in analysing efficiency of Korean banks, Hao, et al. (2001) concluded that the financial liberalisation measures of 1991 did not spur improved cost efficiency in the banking sector. Instead, better performance was driven by asset growth and low expense ratios.

In the Indian sub-continent, results show that overall efficiency was found to be higher for public sector banks than for foreign and domestic private banks (Ataullah & Le, 2006; Sensarma, 2006; Ataullah, Cockerel, & Le, 2004; Shanmugan & Das, 2004). However, the results of di Patti and Hardy (2005) for Pakistan show that efficiency gains were high for all banks during the early years of privatisation but subsequently, the level of efficiency for privatised banks decreased. New private domestic banks were the shining example of better performance while foreign banks lagged behind. For Sri Lanka, Seelanatha (2007) found regressive efficiency performance of different forms for the banks during the first half of the study period. By the end of the second half, banks had marginally recouped efficiency losses. The author attributes the declining trend in efficiency scores in the earlier period to the short-term policy effects and argues that in the long-run, banks in Sri Lanka should record improvement in efficiency as reforms get entrenched and the macroeconomic environment becomes more stable.

From the review above, the general observation is that in developed countries and emerging economies, reforms were accompanied by an improvement in bank efficiency. The analysis also shows that more efficiency gains could accrue from further development in the financial sector. However, some studies reported of dampened efficiency performance. Berger and Humphrey (1997) submit that deregulation may not always improve efficiency and productivity due to other intervening incentives especially in the early years of reform. These factors stem from short-term effects related to the instability in the macroeconomic environment.

4.4.2 Empirical literature: sub-Saharan Africa

The literature on the studies of bank efficiency in sub-Saharan African (SSA) countries has lagged those of other developing countries. Only recently has there been a noticeable increase in the number of studies analysing the efficiency performance of commercial banks. The slow pace in growth of the bank efficiency literature for SSA is partly due to lack of bank-level data during the control regimes. Bank efficiency studies for SSA include Ikhide (2000; 2008) and Adongo, et al.

(2005a; 2005b) for Namibia; Hauner and Peiris (2008) and Beck and Hesse (2009) for Uganda and Čihák and Podpiera (2005) for Kenya, Tanzania and Uganda.

In the efficiency analysis of Namibian banks, Ikhide (2000; 2008) and Adongo, et al. (2005a; 2005b) reached contrasting conclusions. The studies by Ikhide (2000; 2008) measure economies of scale and scope with the suggestion that banks in Namibia were characterised by inefficiency. On the other hand, using an alternative profit X-efficiency approach, Adongo, et al. (2005a; 2005b) argue that Namibian banks compared relatively well with international evidence. The variations in conclusions may be due to differences in the approaches used to measure banking efficiency.

For east African countries, Čihák and Podpiera (2005) and Hauner and Peiris (2008) reach similar conclusions and observe that the increase in bank competition due to breaking up of entry barriers and entry into the sector of foreign banks was associated with a rise in efficiency. Of the two studies, Hauner and Peiris (2008) deduced changes in efficiency using competition analysis while Čihák and Podpiera (2005) applied descriptive analysis based on several bank performance indicators, including interest margins. Recent evidence by Beck & Hesse (2009) shows that the high spreads characteristic of Ugandan banks is suggestive of inefficiency performance in the banking industry. It is important to point out that the accounting ratios and narrow measures of efficiency such as spreads are only indicative of actual bank performance and may not provide reliable estimates of banking efficiency (World Bank, 2006).

More recently there have been studies conducted for a group of African countries. For example, Chen (2009) estimated a frontier function for SSA middle-income countries and evaluates the determinants of efficiency. The study shows that bank inefficiency was in the order of 20-30 percent. Foreign banks were found to be more efficient than public and domestic private banks. Chen (2009) also argues that besides macroeconomic stability, bank competition and financial development, institutional factors also explain the differences in bank efficiency for the sample countries.

Kablan (2007) and Kirkpatrick, et al. (2008) also provide evidence of banking efficiency for countries in the West African Monetary Union (WAEMEU) and SSA, respectively. Kablan (2007) observed that cost efficiency was generally higher in the majority of WAEMU states except for Burkina Faso and Cote D’Ivoire while Kirkipatrick, et al. (2008) found lower profit X-inefficiency than cost X-inefficiency for SSA countries. Both studies observed that inefficiency was sensitive to the quality of loans and bank soundness. Interestingly, Kirkipatrick, et al. (2008) also found a negative effect of financial liberalisation but submit that foreign bank penetration helped improve cost X-efficiency. Although many countries were already implementing financial reforms during this period, it is important to note that the banking industry in a majority of these countries also experienced significant macroeconomic instability. It is not surprising therefore that cost-efficiency was susceptible to risk and solvency factors and the turbulent economic environment.

Figueira, et al. (2006) interrogated the effect of ownership structure on efficiency of African banks, using a much larger sample. Using both the parametric SFA and non-parametric DEA, they show that private domestic banks did not appear to outperform public banks. However, foreign owned banks posted higher efficiency than any of the two bank groups, namely domestic private and public banks. Efficiency indicators also showed wide variations across different bank categories.

4.4.3 Key findings from the empirical literature

The review of literature highlights growing importance of studies in developing countries, spurred by availability of bank level data and the need to assess the efficacy of banking sector reforms. In general, the studies show that banks could still improve the level of efficiency by cutting excess costs. Nonetheless, there is ample evidence to suggest that banking efficiency improves with the breadth and depth of liberalisation and deregulation policies. The reported reverse effect in a few studies is due to the short-term effects of liberalisation such as credit rationing, high spreads and weakening loan quality. These problems tend to be exacerbated under an unstable macroeconomic environment which is often associated with early years of reforms.

Indeed, Denizer, et al. (2007) have observed that a number of developing countries experienced instability in the macroeconomic environment characterised by high inflation, slower economic growth and other economic ills, including banking crises. These factors tend to produce a distorted incentive structure for banks, making resource allocation and utilisation to achieve greater efficiency a much harder task.

The determinants of banking efficiency are all encompassing. They include bank-specific, market structure, macroeconomic and institutional factors. The type of ownership structure is also an important source of bank efficiency. However, across a range of studies, there is less conclusive evidence on whether foreign banks are more efficient than their public and domestic counterparts. In general, foreign banks perform better when institutions are of good quality in host countries.

4.5 Methodological Framework

4.5.1 Choice of appropriate frontier model

There is no consensus on the preferred frontier method to apply in benchmarking efficiency performance. Berger and Humphrey (1997) reported that there are approximately an equal number of studies using non-parametric methods and applications of parametric methods in analysing efficiency of financial institutions. Yet, other studies have estimated efficiency using both approaches (Beccalli, Casu, & Girardone, 2006; Resti, 1997; Weill, 2004). According to Resti (1997) there is no significant difference between the two frontier approaches. When differences occur, they can be explained by revisiting the intrinsic features of the models. On the other hand, Eisenbeis, et al. (1999) argue that efficiency scores derived from the DEA are two to three times larger than those estimated by the SFA. However, the patterns of scores across banks are similar and there is a relatively high correlation between the efficiency scores derived from the two methods.

Against the above background, there is no loss of generality in using either methodology to analyse efficiency in banking. Consequently, we estimate bank efficiency using the stochastic frontier approach. The SFA allows for simultaneous estimation of the cost function and the determinants of cost efficiency in the context of Battese and Coelli (1995). The DEA does not support this estimation procedure. A further advantage of the SFA over the DEA is that with panel data, it is easier to handle problems assosiated with statistical noise in the sense that additional information from multiple time periods is incorporated into the estimation process. The SFA has also received wide application in efficiency studies for SSA. Therefore, we are able to compare our results with previous research based on similar methodology. However, the usual caveats about potential problems arising from misspecification of the cost function remain.

4.5.2 Formulating a stochastic frontier cost function

In order to remain relevant with economic theory, we postulate that overall cost efficiency is a precondition for bank performance and for regulatory policy analysis. In accordance with Bauer, et al. (1998), we conjecture that Zambian banks minimise costs subject to a specific technology. Due to institutional and regulatory factors, commercial banks also face policy and institutional constraints which impose additional costs on production optimisation decisions, thus affecting efficiency performance. Therefore, these constraints form part of the decision variables for individual banks.

Accordingly, following standard literature, see for instance, Bikker and Bos (2005) our hypothesised cost function is specified as:

$$TC(\boldsymbol{w},\boldsymbol{y}) = \min \boldsymbol{w}.\boldsymbol{x} \; s.t. \, T(\boldsymbol{x},\boldsymbol{y},\psi) \tag{4.1}$$

where $TC(\boldsymbol{w},\boldsymbol{y})$ is the cost function, TC is total economic costs incurred by the bank, $\boldsymbol{y}$ is a vector of bank output/services, $\boldsymbol{w}$ is a vector of input prices, $\boldsymbol{x}$ is a vector of factors of production, $T(\boldsymbol{x},\boldsymbol{y},\psi)$ is the transformation function and ψ denotes a vector of bank specific and other factors which shift the bank's cost function.

Equation (4.1) can be used to solve for cost minimisation conditional input demand functions by invoking the Envelop theorem and applying the Shepherd's Lemma. Forming the Lagrangian function of Equation (4.1) and differentiating it with respect to input prices, we can solve for the equilibrium input demand functions as given by:

$$x_j^* = x_j^*\left(w_j, y_q, \psi_S\right), \;\; j = 1,2,...,J; \;\; q = 1,2,...,Q; \;\; s = 1,2,...,S \tag{4.2}$$

where y_q is qth measure of bank output, w_j is the jth input price for the individual bank and ψ_s denotes sth cost shifters. An asterisk denotes optimal input demands. Note that Equation (4.2) states that equilibrium conditional input demand is a function of the level of input prices, output and other factors which affect the bank's production process. Substituting (4.2) into (4.1) yields a firm's minimum cost function expressed as

$$TC = w_j x_j^*\left(w_j, y_q, \psi_s\right) = TC\left(w_j, y_q, \psi_s\right) \tag{4.3}$$

The intuition behind Equation (4.3) is that banks minimise costs by choosing an optimal number of inputs at given prices, a predetermined level of output sold at market prices and bank-specific and exogenous factors which affect its level of production in a given period.

4.5.3 The econometric specification and modelling strategy

The cost function given by Equation (4.3) embodies a long-run equilibrium relationship between costs on one hand and inputs, outputs and exogenous factors on the other, and it possesses all the properties of a cost function. In estimating the equation, care must be taken in selecting time periods that are marked by (relative) stability. For the Zambian case under consideration, this requirement is satisfied since the study relates to the period after the banking crisis, in which banks have continued to exhibit relative stability.

Derivation of the minimum cost function depicted above is a useful starting point in formulating the cost efficient frontier along the lines of Aigner, et al. (1977) and Meesen and van den Broeck (1977). Battese and Coelli (1995) posit that a similar approach can be extended to the analysis of efficiency in the panel data setup. Thus, consistent with Battese and Coelli (1995), Equation (4.3) is reformulated yielding the stochastic frontier cost function expressed in general logarithmic form as:

$$\ln\left(TC_{it}\right) = f\left(\ln\left(y_{qit}, w_{jit}, \psi_{sit}\right), \beta\right) + \omega_{it}; \qquad \omega_{it} = v_{it} + u_{it} \qquad (4.4)$$

where ω_{it} is the composite error term for the ith bank $\left(i = 1,2,.......,N\right)$. Equation (4.4) is a stochastic frontier cost equation where TC, y, w and ψ are as defined before, β is a vector of parameters to be estimated. According to Aigner, et al. (1977), v_{it} is a random uncontrollable error term assumed to be standard normal, independently and identically distributed with zero mean and constant variance, i.e., $v_{it} \sim N\left(0, \sigma_v^2\right)$. Furthermore, $u_{it} > 0$ is a one sided controllable component of the disturbance term, capturing cost inefficiency, that is, deviation of actual cost from the minimum possible level of costs given the hypothesised frontier. Put differently, it reflects the inability of bank i at observation t to attain the minimum cost level defined by the 'best practice' frontier. The distributional asusmption of u_{it} is explained below.

In line with Battese and Coelli (1995) the panel data inefficiency conditional model is given by

$$u_{it} = \theta_0 + \sum_n \theta_{n,it} z_{n,it} + \varepsilon_{it} \qquad (4.5).$$

Equation (4.5) specifies the cost inefficiency of each bank where z_{it} are bank specific and other exogenous variables that determine cost inefficiency and the θs represent the associated coefficients depicting the statistical relationship between cost inefficiency and its possible determinants. Finally, ε_{it} is a collection of the effects of other unobserved systematic differences across banks that may not have been fully incorporated into the model and is defined by the truncation of the normal distribution with zero mean and constant variance, that is, $\varepsilon_{it} \sim N(0, \sigma_\varepsilon)$.[23] Equation (4.5) allows us to estimate time-varying inefficiency scores which can then be related to factors that drive efficiency performance.

There are no a priori reasons to prefer a specific distribution of the inefficiency term (u_{it}). Here we assume that u_{it} is assumed to be truncated-normal, first introduced by Stevenson (1980). With a truncated-normal distributed disturbance, the one-sided error term is taken to be the variable obtained by truncating at zero the distribution of a variable with a possibly non-zero mean (Fu & Heffernan, 2007). The truncated-normal distribution also provides more flexible representations of the pattern of efficiency in the data. Therefore, following Jondrow, et al. (1982), the firm specific inefficiency term, u_{it} is decomposed using the conditional mean distribution approach for a truncated normal distribution as shown below:

$$E(u_{it} \mid \omega_{it}) = \frac{\sigma\gamma}{1+\gamma^2}\left[\frac{\phi(\mu^*/\sigma)}{\Phi(\mu^*/\sigma)} + \frac{\mu^*}{\sigma}\right] \qquad (4.6)$$

where $E(u_{it} \mid \omega_{it})$ is an unbiased though inconsistent estimator of u_{it}, $\gamma = \sigma_u / (\sigma_v + \sigma_u)$ measures the relative importance of u_{it} and v_{it} to the composite error term, ω_{it} and must lie between 0 and 1

[23] According to Battese and Coelli (1995) z_{nit} and ψ_{sit} may be the same variables in that factors that affect banks' costs may also influence its cost efficiency performance.

(Battese & Corra, 1977). As $\gamma \to 0$ the effect of the noise error term, v_{it} dominates the stochastic cost inefficiency error term u_{it} in determining the variation of the global residual ω_{it}. Conversely, when $\gamma \to 1$ the cost inefficiency effects are stronger in driving the composite error term, ω_{it}, implying that the stochastic frontier is the correct model specification. The terms Φ and ϕ are respectively, the standard normal cumulative density function and standard normal density function evaluated at μ^{*}/σ where $\mu^{*} = (\omega_{it}\gamma)/\sigma + u/(\sigma\gamma)$. Once cost inefficiency estimates in Equation (4.6) are obtained, the time variant bank specific cost inefficiency index (CI) is calculated by[24]

$$CI_{it} = E\left[exp\left(u_{it}^{*}\right) \mid \omega_{\iota\tau}\right] \tag{4.7}$$

To estimate the cost function depicted by Equation (4.4) we adopt a transcendental logarithmic (translog) cost formulation first proposed by Christensen, et al. (1973). This specification enables us to estimate cost inefficiency and examine the relationship between cost inefficiency and factors that affect it. Therefore, following Battese and Coelli (1995) we estimate the translog frontier cost function and the inefficiency model in a single-step using maximum likelihood estimation (MLE) procedure. The alternative is a two-step procedure espoused in other studies.[25, 26] It should be noted that the translog function is a more general specification and incorporates the more restrictive Cobb-Douglas function. It is therefore suitable for frontier estimation. Accordingly, Equation (4.8) represents the empirical translog cost function:

---

[24] The converse is cost efficiency given by the inverse of CI. In the empirical section below, we discuss the evolution of cost efficiency, depicted as $CE_{it} = E\left[exp\left(-u_{it}^{*}\right) \mid \omega_{\iota\tau}\right]$.

[25] We argue that it is inappropriate to use the two-step regression procedure because it contradicts the assumption of identically distributed inefficiency effects in the stochastic frontier model and because of possible misspecification of the cost frontier; this could introduce some bias, see Battese and Coelli (1995). Therefore, one can avoid this problem by simultaneously estimating the frontier function and the inefficiency model.

[26] In studies that employ the two-step procedure, see for instance Kirkpatrick, et al (2008) among others, the translog cost function is estimated jointly with input share equations. However, the Battese-Coelli framework adopted in this analysis abstracts from this estimation approach but allows the researcher to simultanously estimate the translog cost function and the cost inefficiency model, relating cost and inefficiency effects possible explanatory factors. Therefore, it is a more robust way of obtaining estimates of cost efficiency.

$$
\begin{aligned}
\ln TC_{it} &= \beta_0 + \beta_y \ln Y_{it} + \frac{1}{2}\beta_{yy}\left(\ln Y_{it}\right)^2 + \beta_1 \ln w_{Lit} + \beta_2 \ln w_{Fit} + \beta_3 \ln w_{Kit} \\
&\quad + \beta_{11}\left(\ln w_{Lit}\right)^2 + \beta_{22}\left(\ln w_{Fit}\right)^2 + \beta_{33}\left(\ln w_{Kit}\right)^2 + \beta_{12} \ln w_{Lit} \ln w_{Fit} \\
&\quad + \beta_{13} \ln w_{Lit} \ln w_{Kit} + \beta_{23} \ln w_{Fit} \ln w_{Kit} + \sum_{j=L,F,K} \delta_{yj} \ln Y_{it} \ln w_{jit} \\
&\quad + \theta_t t + \theta_{tt} t^2 + \sum_{j=L,F,K} \theta_{tj} t \ln w_{jit} + \theta_{ty} t \ln Y_{it} + \psi_1 \ln\left(BRANCH_{it}\right) \\
&\quad + \psi_2 \ln\left(RISK_{it}\right) + \psi_3 \ln\left(INTERMED_{it}\right) + \omega_{it}, \\
& i = 1,\ldots,N;\; t = 1,\ldots,T
\end{aligned}
\tag{4.8}
$$

where TC_{it} denotes total operating costs (overheard and interest costs), $w_{Lit}, w_{Kit}, w_{Fit}$ are, respectively, unit labour cost, unit cost of capital and unit cost of funds; Y_{it} is total bank output[27], N is the number of banks in the sample, T is the number of quarterly observations per bank and it varies across institutions, $\ln$ is the natural logarithm operator and ω_{it} is as defined above. A time trend, t and its quadratic term, t^2 are included to capture the dynamic changes in costs.

A well behaved cost function must reflect homogeneity of degree one in input prices, implying that for a fixed level of output, total costs must increase proportionally when all factor prices are increased in the same proportion. For this purpose, homogeneity of degree one in input prices is imposed by normalising total costs and input prices by the input price of funds, w_{Fit} before log transformations are undertaken.[28] A similar approach has been used by Weill (2007) and is standard in the literature. Therefore, the following symmetry and parameter restrictions are imposed on the translog cost function:

[27] Although commercial banks are treated as multioutput producing firms, we adopt a single output measure in this study, proxied by total assets, on the basis that the flow of banking services is proportional to its stock of assets. Furthermore, using total assets to depict banks' output ensures that on-balance sheet outputs and off-balance asset items are all captured to avoid understating bank production. This is consistent with overall bank behaviour. Moreover, Zambian banks do not report disaggregated costs for individual assets.

[28] See Cebenoyan (1988) and Zardkoohi, et al. (1986) and references cited therein for details on appropriate forms of normalisations.

$$
\begin{aligned}
&\beta_1 + \beta_2 + \beta_3 = 1; \\
&\beta_{11} + \beta_{12} + \beta_{13} = 0; \\
&\beta_{21} + \beta_{22} + \beta_{23} = 0; \\
&\beta_{31} + \beta_{32} + \beta_{33} = 0; \\
&\delta_{yL} + \delta_{yF} + \delta_{yK} = 0 \quad \text{and} \\
&\beta_{12} = \beta_{21}; \beta_{13} = \beta_{31}; \beta_{23} = \beta_{32}
\end{aligned}
\tag{4.9}
$$

We have also included $BRANCH$, $RISK$ and $INTERMED$ as control variables in the translog cost function. The variable $BRANCH$ is defined as the number of branches operated by commercial banks. It controls for the scale of operation and network density on costs.[29] Therefore, by including $BRANCH$ in the cost function we postulate that the banks' production technology differs in a significant way due to variations in size and other unmeasured factors associated with maintaining the bank branches. We expect $BRANCH$ to carry a positive sign indicating that branch expansion raises operating costs. The variable $RISK$ is the ratio of non-performing loans to total loans. Nonperforming loans (NPLs) are treated as banks' undesirable outputs or costs which decrease the banks' performance. Therefore, $RISK$ is included to capture the impact of poor output quality on bank costs. We expect a positive sign on the coefficient of $RISK$.

Finally, $INTERMED$, the loan-to-deposit ratio (intermediation ratio), is included to measure the effect of loan and deposit fund utilisation of banks. The flow of deposits into the banking sector determines to a large extent the amount of loans a bank can make in a given period. Furthermore, banks that rely more on deposits to finance assets face a higher funding risk than those that hold a relatively higher proportion of equity capital. Conversely, banks with greater market power would produce high deposits and loans, and thus possess a higher loan-to-deposit ratio and therefore lower production costs. In view of this, the sign on the parameter on $INTERMED$ may be indeterminate, depending on which of the two effects dominates.

[29] In some studies, the log of assets is used to capture bank size. However, this may introduce problems of multicollinearity when output is measured by total assets.

4.5.4 Determinants of cost inefficiency

As argued above, the single-step procedure involves simultaneous estimation of the translog cost function with the inefficiency model. The equivalent of Equation (4.5) above, the specification relating cost inefficiency to its determinants, is given by Equation (4.10):

$$\begin{aligned} u_{it} &= \varphi_0 + \phi_1 \ln\left(RISK_{it}\right) + \varphi_2 \ln\left(CONTASS_{it}\right) + \varphi_3 \ln\left(CAPRATIO_{it}\right) \\ &\quad + \varphi_4 \ln\left(OPPCOST_{it}\right) + \varphi_5 \ln\left(INFLATION_t\right) + \varphi_6 \ln\left(TBR_t\right) + \varphi_7 OWNERSHIP + \varepsilon_{it} \end{aligned} \tag{4.10}$$

where, u is the non-negative cost inefficiency for each bank. Equation (4.10) completes the estimation process of the analysis of bank efficiency and it is estimated simultaneously with the stochastic frontier cost function (Equation 4.8) in a single step by MLE. An important aspect of the single step estimation is the ability to relate bank-specific and other factors to inefficiency effects. Battese and Coelli (1995) also posit that when maximum likelihood estimation is involved, a test must be conducted to verify if inefficiency effects exist in the model. Rejecting the null hypothesis would imply that the stochastic frontier cost function incorporating inefficiency effects is the appropriate specification of the model. Accordingly, both the frontier and inefficiency model could have all or some variables in common as drivers of costs may equally influence banks' inefficiency performance. Therefore we hypothesise the following factors as the ones that affect cost inefficiency of Zambian banks.

The variable $RISK$ captures the effect of default risk on banks' inefficiency. We expect bank credit risk to positively affect cost inefficiency of Zambian banks. The variable $CONTASS$ captures the effect of holdings of government securities vis-à-vis traditional intermediation activities of the banking sector. Thus, this variable measures the degree of disintermediation on bank efficiency. According to Hauner (2008), credit to government undermines bank performance in developing countries. Therefore, a positive coefficient is expected for $CONTASS$ indicating that a higher proportion of Treasury securities relative to loans is a source of cost inefficiency.

Changes in the regulatory environment meant to safeguard the banking system may produce some unintended adverse effects on efficiency outcomes since they are products of a convolution of forces. To test this conjecture, two measures of regulatory intensity are used. The variable

OPPCOST captures the impact of the opportunity cost of maintaining cash reserves on banks' cost inefficiency. There is a general understanding that reserve requirements act as an implicit tax on bank performance. Therefore, we expect the coefficient on *OPPCOST* to be positive implying that complying with cash reserve requirements exacerbates the banks' cost inefficiency performance. Another variable, *CAPRATIO* also measures the banks' efficiency responses to regulatory capital intensity. The coefficient of this variable is expected to be negative, indicating that well capitalised banks are also less inefficient because, as Altunbas, et al. (2000) have argued, banks with low inefficiency will have more profits and will be able to retain high proportion of earnings necessary for boosting the capital base.

Conventional economic wisdom suggests that foreign bank presence and private ownership of banks have positive influences on bank efficiency performance than do publicly owned banks. For example, Bhaumik and Dimova (2004) argue that even when faced with a level playing field in terms of regulatory policy, public managers experience less intense performance pressures than their private peers. However, the authors argue further that when competition is intense, ownership structure matters less in performance analysis. On the other hand, when analysing banking efficiency based on public choice theory, Sarkar, et al. (1998) argued that the existence of performance pressures compel private bank managers to adopt better strategies that reduce costs, raise productivity and increase cost efficiency. To investigate the effect of ownership structure on efficiency performance, a dummy variable, *OWNERSHIP* has been included.

Cost efficiency may also be influenced by the macroeconomic environment and monetary policy stance. These factors are captured by the rate of consumer price inflation $(INFLATION)$ and the Treasury bill rate (TBR). Both these variables are expected to carry positve signs. Details on the sample size and description of all variables used in the estimation of the frontier cost function and the inefficiency model are discussed later.

4.5.5 Economies of scale and technological change in banking

In the preceding sections, focus has been devoted to analysing the efficiency of Zambian banks and factors that explain cost inefficiency. This section extends the analysis to presence or lack of economies of scale and technological change in the Zambian commercial banking industry.

Knowledge of scale economies is important for regulatory purposes. Regulatory authorities use estimates of economies of scale to determine the appropriate regulatory policy in respect of mergers and acquisitions and in predicting future industry structure. Recall that cost inefficiency in Zambian commercial banks depicts overutilization (mismanagement) of input resources in the production process. However, it is not clear whether there is some level of output at which banks are said to be enjoying global economies of scale or indeed if cost inefficiency could be ascribed to existence of diseconomies of scale.

Following Mester (1996), estimates of overall economies of scale are obtained by differentiating the translog cost function with respect to output and are evaluated at the mean output and input prices. This yields equation (4.11):

$$ES = \left[\sum \frac{\partial \ln TC_{it}}{\partial \ln Y_{it}} \right]^{-1} = [\beta_y + \beta_{yy} \ln Y_{it} + \sum_{j=L,K} \delta_{jy} \ln w_j + \theta_{ty} t]^{-1} \qquad (4.11).$$

Economies of scale (ES), as denoted by (4.11) measure the relative change in the bank's total costs for a given proportional change in the measured level of output. If ES is greater than, less than or equal to one banks are said to be experiencing economies of scale, diseconomies of scale or constant scale economies, respectively.

Another important concept in banking production is technological progress (TP) which estimates a relative change in operating costs due to passage of time. This might occur when the banks' investment in some technologically improving production techniques in turn allows them to capture efficiency gains from a reduction in production costs. Technological change is therefore an encompassing concept that includes financial innovations, changes in competition intensity and the regulatory environment and dynamics in credit allocation process. Accordingly, technological change may impact on the banks' provision of financial services and the instruments used to provide these services. For example, technological change could be underpinned by the spread of automated teller machines (ATMs) and adoption of better risk-management techniques.

Equation (4.12) below provides an approximation of technological progress in Zambian banking assuming technological change directly affects the banks' cost function. Put differently, we conjecture that banks have been subjected to the same form of technological shocks over the sample period.

$$TP = -\sum \frac{\partial \ln TC_{it}}{\partial t} = -\left[\theta_t + \theta_{tt} t + \sum_{j=L,K} \theta_{tj} \ln w_j + \theta_{ty} \ln Y_{it} \right] \quad (4.12)$$

The novelty of this specification is that it does not require knowledge of the timing of the sources of technological changes, given that it is hard to pinpoint when banks adopt those technologies and at what point effects on the cost structure take hold. Since we are using a cost function, technological progress occurs when $TP < 0$ meaning that costs decrease with technological advancement and when $TP > 0$ technological change is regressive depicting increasing production costs as banks adopt new technological means of production. From (4.12), pure technological change, that is, the decline in costs keeping constant input proportions, is given by the first two elements in the square brackets on the right hand side. The third term denotes non-neutral technical change associated with the effect of input factor prices. Finally, the last term measures scale augmented technological progress, related to the effect of quantity of bank outputs on costs.

4.5.6 Data and variable description

In order to analyse bank cost inefficiency, we employ a uniquely assembled quarterly bank level data set. All 15 Zambian banks continuously operating during the period January 1998 to December 2006 form the population of this study. This period is especially special in the Zambian banking industry because it follows changes to the supervisory and regulatory environment and enactment of banking legislation in response to the banking crisis of the mid-1990s. This period was also characterised by episodes of relative macroeconomic stability and rebound in economic growth especially after 2002. It is expected that these economic dynamics would have altered the banks' behaviour in a significant manner and should be reflected in efficiency performance.

All the data used in the analysis were gleaned from the commercial banks' monthly consolidated balance sheets and income statements submitted to the Bank of Zambia. Therefore, the information provides all inputs for the necessary calculations. Supplementary data from the fortnightly statistics and data maintained by the Financial Markets Department of the Bank of Zambia under the AREMOS data project were also used as sources. Details on the definitions and descriptive statistics of the variables used in the analysis are presented in Table 4.1 below.

Table 4-1: Variables used in estimating cost efficiency and determinants of inefficiency

Variable Symbol	Variable Name	Description and measurement	Mean	Median	Std Dev.
TC	Total costs	Interest expenses plus operating costs	52,200.450	54,027.880	20,222.870
Y	Bank output	Stock of total assets	329,127.700	137,957.000	402,073.500
w_L	Unit price of labour	Total personnel expenses expressed as a proportion of total assets	0.005	0.004	0.004
w_F	Unit price of funds	Total interest expenses on deposits and other borrowed funds divided by total deposits and borrowed funds	0.014	0.006	0.122
w_K	Unit price of physical capital	Sum of all other expenses (on building, equipment, furniture, etc.) divided by stock of fixed and other assets	0.110	0.054	0.410
BRANCH	Branches	Total number of bank branches operated by an individual bank per given period	11	7	13
RISK	Portfolio credit risk	Nonperforming loans expressed as a proportion of total loans	0.092	0.074	0.107
INTERMED	Intermediation ratio	Proportion of total loans to total deposits (Loan-to-deposit ratio)	12.239	0.408	35.477
CONTASS	Asset concentration ratio	Proportion of total government securities to total loans (proxy for disintermediation)	2.748	0.836	8.405
INFLATION	Inflation rate	Changes in consumer price index (CPI), percent per annum, expressed on a quarterly basis	20.871	20.600	5.846
TBR	Treasury bill rate	Yield rate on 91-day Treasury bill paper (percent per annum) expressed on a quarterly basis	27.301	32.400	12.507
OPPCOST	Opportunity cost of cash statutory reserves	Interest foregone on cash reserve requirements expressed as a proportion of interest expenses	695.853	213.556	1065.366
CAPRATIO	Regulatory capital adequacy ratio	Banks' total capital (Tier I and Tier II capital) expressed as a proportion of risk-weighted assets	0.427	0.300	0.448
OWNERSHIP	Ownership dummy	Dummy variable for bank ownership structure (1 for foreign and domestic private banks, zero for public banks)			

Note: Total costs and bank output (assets) are measured in millions of current Zambian Kwacha (K'million).

To calculated interest foregone, the risk free 91-day Treasury bill rate is applied to total cash reserves on the assumption that under a regime of zero reserve requirements, banks would invest their funds in less risk assets for a guaranteed income stream. Of course, banks can also invest any freed resources in other assets, including foreign exchange deposits abroad and loans. However, these assets are subject to intertemporal uncertainty during the period of investment. Securities are less prone to uncertainty.

Source: Bank of Zambia (BoZ)

4.6 Empirical results and discussion

The ensuing discussion reports results obtained from the translog stochastic frontier cost function and the accompanying inefficiency model. First we present econometric tests on the suitability of the frontier and inefficiency effects model specification. Secondly, we discuss empirical results of the translog frontier cost function and estimates of the inefficiency model. A discussion on the properties of scale economies and technological change follows afterwards.

4.6.1 Estimation results of the translog stochastic frontier cost function

The estimation of the stochastic frontier cost function was performed using maximum likelihood function incorporated into Stata 10.0 following the parameterisation by Battese and Coelli (1995). As noted earlier, the Battese-Coelli approach provides parameter estimates for the cost function as well as for determinants of cost inefficiency assuming a more general truncated-normal distribution of the error component. The truncated-normal distribution of the inefficiency term nests the half-normal distribution as a special case (Fujii, 2001).

Thus, as a robustness check, a more restrictive half-normal distribution of the inefficiency effects and a specification assuming heteroscedasticity (Bottasso & Sembenelli, 2004; Hadri, 1999) were estimated but results were unreliable due to model instability. Estimations based on the two specifications failed to converge even at a higher number (>5000) of iterations. The failure of the heteroscedasticity model may be attributed to a flat log likelihood which is caused by insufficient degrees of freedom. On the other hand, convergence in the Battese and Coelli (1995) specification was achieved after only 33 iterations.

Table 4.2 below gives the empirical results of the maximum likelihood parameter estimates obtained from the normalised translog frontier cost function. Overall the translog cost function is well behaved and passes a battery of diagnostic tests. The null hypothesis that a restrictive Cobb-Douglas cost function would have been the appropriate specification is rejected at 1 percent. The test statistic is asymptotically distributed as a χ^2 variable with degrees of freedom given by the number of imposed restrictions. Therefore, the test exonerates the generalised translog cost function as the appropriate specification.

Table 4-2: Empirical results of the stochastic frontier translog cost function

	Coefficient	Parameter	t-statistic	p-value
Intercept	β_0	4.448	7.573	0.000***
ln*(w_L)*	β_1	0.953	7.619	0.000***
ln*(w_K)*	β_3	0.182	1.397	0.162
ln*(tass)*	β_y	0.310	2.679	0.007***
*1/2 (*ln*(tass))*^{2}	β_{yy}	0.050	4.211	0.000***
*1/2 (*ln*(w_L)*^{2}	δ_1	0.068	2.350	0.019**
*1/2 (*ln*(w_K))*^{2}	δ_3	0.110	3.425	0.001***
ln*(w_L)*ln*(w_K)*	δ_{13}	-0.081	-2.893	0.004***
ln*(w_L)*ln*(tass)*	δ_{1y}	-0.008	0.623	0533
ln*(w_K)*ln*(tass)*	δ_{3y}	-0.007	0.475	0.634
t	θ_t	0.022	2.411	0.016**
t^2	θ_{tt}	-0.001	-3.245	0.000***
t ln*(tass)*	θ_{ty}	-0.0001	0.088	0.930
t ln*(w_L)*	θ_{tL}	0.001	-0.722	0.470
t ln*(w_K)*	θ_{tK}	-0.003	-2.827	0.005***
Control Variables				
ln*(BRANCH)*	ψ_1	0.120	11.011	0.000***
ln*(RISK)*	ψ_2	-0.015	1.740	0.082*
ln*(INTERMED)*	ψ_3	-0.032	-2.759	0.006***
Diagnostics				
Log likelihood function		308.812		
Wald chi square (p-value)		0.000***		
σ_u^2		0.099		
σ_v^2		0.006		
σ^2		0.106		
γ		0.940		
LR test of one-sided error		143.742***		
LR test of Cobb-Douglas functional form		103.710***		
Observations		380		

Significance level: * p<0.10, ** p<0.05, *** p<0.01
Source: Author's own calculations based on BoZ data

More importantly, the test for the one-sided inefficiency error as the dominant structure cannot be rejected at 1 percent. The likelihood ratio (LR) statistic was calculated as 143.742 against the critical value of 42.27.[30] This test result is further reinforced by the fact that $\gamma = 0.94$ was found to be statistically significant at 1 percent level, indicating that the inefficiency term dominates the random

[30] Appropriate critical values for the LR test of the one-sided error component which follows a mixed chi square distribution are provided in Table I of Kodde and Palm (1986).

error term in the overall error variance, σ^2. This result also means that we cannot use ordinary least squares (OLS) to estimate cost inefficiency since it disregards the bank-specific inefficiency component. Thus, deviations from the best-practice frontier are largely driven by bank-specific inefficiency effects.

Sample parameter estimates are plausible, consistent with apriori expectations. We observe a positive and significant coefficient on the normalised unit price of labour. The estimated coefficient shows that a unit increase in the normalised labour factor price directly translates into 0.95 percent increase in total costs. This means there is a near correspondence between labour costs and overall bank expenses. Conversely, the estimated coefficient for the normalised unit price of capital was found to be statistically insignificant, despite carrying a positive sign. Since equipment and buildings are of a fixed nature, costs related to these assets tend to be of less significance over time. Worthington (1998) also found an insignificant coefficient for the normalised price of physical capital for the Australian building societies. A similar result was reported by Ansari (2007) for the Pakistan banking sector. Fu and Heffernan (2007) found a negative and significant parameter estimate on the unit price of physical capital for the Chinese banking sector.

Estimates for the translog frontier model also depict an important effect of bank output on costs. The point estimates for the output measure and its quadratic term are statistically significant at 1 percent. This shows that bank costs increase with the scale of production. The results of the time trend show that banks' costs increase with time in a contemporaneous fashion, but with passage of time, costs decrease, allowing banks to obtain efficiency gains through 'learning by doing' effects. Thus, a positive parameter estimate for the time variable is observed but the quadratic term carries a negative sign which may be indicative of narrowing cost inefficiencies.

Control variables in the frontier cost specification are employed to isolate the effect of some of the important bank-specific factors on banks' costs. From the estimated results we observe a positive coefficient on the log of branch network density. This finding underpins that obtained for the output measure (total assets), indicating that adequately controlling for differences in the scale of operation produces robust results. For this variable, a percent increase in branches raises operating costs by 0.12 percent implying that maintaining a large branch network has some cost implications. According to Giokas (2008), banks strive to provide services of better quality within reach of their customers by expanding branch network. However, this also leads to an increase in operating costs of bank branches. Similarly, Grigorian and Manole (2006) argue that in highly inflationary economic conditions, banks tend to increase the number of branches but this raises overhead costs thereby weakening efficient performance. These views reinforce the argument by Berger, et al. (1997) that although over branching raises revenues from providing extra customer convenience, this comes at a cost which is reflected in disproportionately high X-inefficiencies.

The above arguments also apply to the Zambian banking sector. For a long time, Zambian banks have operated under an environment characterised by high rates of inflation. Furthermore, over the sample period, the number of branches increased, as banks reached out to customers to gain a competitive edge over peers. However, costs also increased, confirming the positive effect of branch network density on costs. In order to cut fixed costs related to branching activity, banks should adopt better technologies such as increased use of automated teller machines (ATMs) and telephone and internet banking and rely less on brick and mortar. Although the initial capital outlay of setting up such technological infrastructure is high, these costs tend to diminish over time thereby enabling banks to realise significant cost savings. According to Boitumelo and Valadkhani (2008), adoption of self-service technologies can lead to a substantial reduction in the service delivery costs.

The converse is observed for the intermediation ratio and risk factor effects. These two impact bank costs negatively. For the liquidity ratio, this effect is significant at 1 percent level but accounts for only 0.03 percent of the reduction in bank costs. Nonetheless, it indicates that banks with a higher proportion of loans-to-deposits are more able to contain funding costs resulting in moderate cost savings. The impact of nonperforming loans is less robust with a coefficient of 0.02 percent and only statistically significant at the 10 percent level.

4.6.2 Determinants of cost inefficiency

Table 4.3 below summarises results of the multivariate conditional mean (CM) inefficiency model (Equation 4.10), jointly estimated with the stochastic frontier cost function (Equation 4.8). In this model, the estimated inefficiency effects were regressed on a set of variables which affect cost inefficiency. The variables capture the impact of regulation, risk, macroeconomic policy environment and ownership structure on banks' cost inefficiency. From Table 4.3 the null hypothesis that $\gamma = \varphi_0 = \varphi_1 = \varphi_2 = \varphi_3 = \varphi_4 = \varphi_5 = \varphi_6 = \varphi_7 = 0$ was rejected at 1 percent as shown by the LR test of one-sided error component (see above for appropriate critical values). This implies that Zambian banks can be modelled as being driven by inefficiency effects.

Table 4-3: Parameter estimates of the conditional mean (CM) inefficiency model

	Coefficient	Parameter	t-ratio	p-value
Intercept	φ_0	9.729	4.457	0.000***
$\ln(RISK)$	φ_1	0.443	2.151	0.031**
$\ln(CONTASS)$	φ_2	0.488	2.926	0.003***
$\ln(CAPRATIO)$	φ_3	-0.662	-2.760	0.006***
$\ln(OPPCOST)$	φ_4	-0.331	-2.920	0.004***
$\ln(INFLATION)$	φ_5	0.874	1.963	0.050**
$\ln(TBR)$	φ_6	-0.397	-2.089	0.037**
$OWNERSHIP$	φ_7	0.666	2.180	0.029**
Diagnostics				
σ_u^2		0.071		
σ_v^2		0.006		
σ^2		0.077		
γ		0.923		
LR test of one-sided error		143.742***		
Observations		380		

Significance level: * p<0.10, ** p<0.05, *** p<0.01
Source: Author's own calculations based on BoZ data

Empirical results also show that both measures of regulatory pressure, namely capital ratio $(CAPRATIO)$ and opportunity cost of cash reserves $(OPPCOST)$ carry negative signs and are significantly different from zero at 1 percent. For the capital ratio, the estimated coefficient is consistent with bank behaviour and shows that adequately capitalised banks have higher cost efficiency and may be subject to less regulatory pressure than poorly capitalised banks. For a unit increase in regulatory capital ratio, banks recorded two thirds of a percent reduction in cost inefficiency. This corroborates previous research findings, which have established that regulatory capital should be inversely related to inefficiency on grounds that banks with low inefficiency will have higher profits. Hence, they are able to use these profits to bolster their capital buffer which serves as an indicator of good reputation (Altunbas, Liu, Molyneux, & Seth, 2000; Berger & De Young, 1997; Carvallo & Kasman, 2005; Fries & Taci, 2005). Therefore, a high capital ratio is associated with better efficiency performance because banks can borrow at low interest rates since they are perceived as less risky.

Turning to the measure of the opportunity cost of reserves, the results show that compliance with cash reserve requirements did not impair the banks' ability to perform efficiently. The coefficient on $OPPCOST$ was estimated at 0.33. Since Zambian banks hold large amounts of excess reserves, this helps them to remain regularly compliant with statutory reserve requirements. Therefore, this result means that there is no opportunity cost explicitly associated with non-remuneration of reserve deposits. Rather, more compliant banks with a high proportion of statutory reserves also recorded better cost efficiency.

The ratio of nonperforming loans-to-gross loans $(RISK)$,the variable capturing output quality is positive and significant at 5 percent level. This means that one percent deterioration in the banks' loan quality results in loss of efficiency performance in the order of 0.44 percent. Similar findings have been reported in previous studies. For example, Carvallo and Kasman (2005) found that high risk undermined banks' ability to improve cost efficiency performance of Latin American and Caribbean banks.

The statistically significant positive coefficient for $CONTASS$ (ratio of securities to total loans) is also expected since higher Treasury bill holdings (credit to government) relative to loans augments banks' cost inefficiency. This could be attributed to the cushioning effect Treasury securities have on banks' operations, which serves as a disincentive to devise ways of curbing managerial and other organisational inadequacies and lapses. Hauner (2008) attributed low efficiency performance of banks in developing countries to provision of credit to government. In a majority of developing countries, credit to government is in form of securities holdings.

The measure of macroeconomic uncertainty and policy stance $(INFLATION)$ is unambiguously positive, indicating that high inflation prevents banks from operating close to the efficient frontier. Clearly, a volatile macroeconomic environment is counterproductive to banks' efficiency improvement. In particular, the magnitude of the estimate on the log of inflation validates calls for macroeconomic stability as a pre-requisite for an efficiently functioning banking system. Recent studies by Chen (2009) for sub-Saharan African middle income countries and Weiss (2007) also showed that inflation is detrimental to banks' cost efficiency performance. Moreover, Yildirim (2002), Denizer, et al. (2007) and Zaim (1995) all argued that uncertainty in the macroeconomic setting slowed down the recovery of Turkish banks' efficiency performance. This effect tends to be strong when banks are emerging from a crisis.

Finally, foreign bank presence in the Zambian banking system does not seem to have improved cost efficiency in the industry. The estimated coefficient on the dummy variable

OWNERSHIP indicates that relative to other forms of bank ownership, foreign banks were two thirds of a percent less efficient. The finding is consistent with de Luna Martinez's (2006) observation that foreign banks have not performed as well as the open policy dictates. It also corroborates findings by Sensarma (2006) that foreign banks operating in India were the worst performers in efficiency terms between 1986 and 2000 compared with state owned and private domestic banks. Lensink, et al. (2008) also found that foreign bank presence negatively affects efficiency, although this effect is less pronounced in countries with good governance institutions. In Zambia, banks operate in a very challenging and informationally opaque environment. The absence of specialised courts to enforce credit contracts further complicates the banks' ability to recover debt from defaulters. Potentially, these factors affected efficiency performance of foreign banks more than they did for other bank categories, especially public owned bank which enjoyed considerable state support in various forms.[31]

However, the result is in contradiction with both theoretical postulations and empirical observations which stipulates that foreign banks infuse efficiency in the host banking sector and serve as a benchmark for improved performance by the public and domestic private banks. The result is also inconsistent with the public choice theory discussed earlier in the context of Sarkar, et al. (1998).

4.6.3 Characteristics of average bank-specific efficiency scores

Table 4.4 below presents a ranking of average bank-specific cost efficiency scores estimated from the translog frontier cost function.[32] For reasons of confidentiality, banks have been coded alphabetically (column 1). The corresponding average efficiency score is given in

[31] It must be noted that in general we would expect lower efficiency performance of public sector banks given their profligacy.

[32] In the Battese-Coelli specification framework, inefficiency scores range between 1 (depicting the most efficient bank) and infinity (depicting the least efficient bank). For comparative purposes we rank banks according to cost efficiency scores, obtained by inverting the Battese-Coelli inefficiency estimates. Cost efficiency ranges between zero and unity. The most efficient bank has a score of one while zero denotes least efficiency score. Depatures from efficiency performance (best practice frontier) are calculated as the difference between 1 and the estimated efficiency score. Appendix II presents efficiency scores per bank per quarterly observation obtained from estimating Equation (4.8).

column 2 and column 5 gives position of the bank's efficiency score relative to others. Therefore, a bank ranked first means that it is more efficient than a bank ranked second, and so on.

Cost efficiency estimates show that for the banking sector as a whole, banks operated below the frontier, with mean cost efficiency of 92.0 percent, depicting cost inefficiency of 8.0 percent. This means that if all banks were utilising the best practice production technology they could have improved their efficiency performance by saving up to 8.0 percent in total costs. For individual banks, the cost efficiency scores indicate that relative to the common best practice frontier, the most efficient bank (Bank D) had a cost efficiency score of 96.2 percent, that is, 4.9 percent more efficient than the average bank in the sample. To move to the efficient frontier, Bank D could have cut its costs by 3.8 percent; this reflects its level of cost inefficiency. Conversely, the least cost efficient Bank M had a cost efficiency score of 69.2 percent depicting input wastage of about 30.8 percent. Thus, bank M could have improved its cost performance by shedding off excess costs that prevented it from moving closer to the efficient frontier.

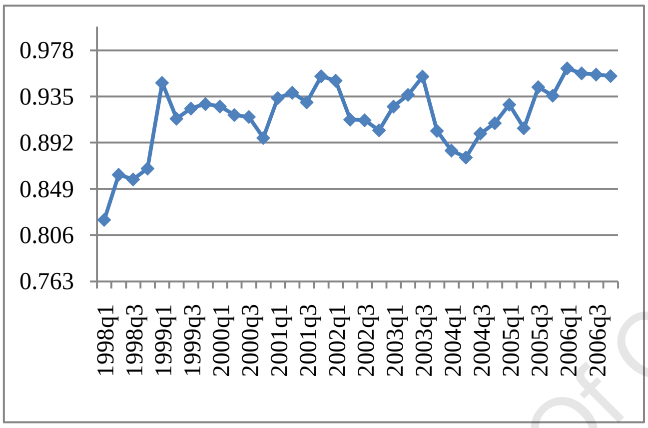
Table 4-4: Mean cost efficiency ranking of Zambian banks

Bank Code	Mean Efficiency	Minimum	Maximum	Efficiency Rank	Ownership
A	0.931	0.688	0.977	10	Foreign
B	0.960	0.904	0.982	3	Foreign
C	0.962	0.921	0.979	2	Foreign
D	0.962	0.934	0.974	1	Foreign
E	0.805	0.693	0.934	13	Foreign
F	0.833	0.538	0.979	12	Foreign
G	0.923	0.773	0.975	11	Foreign
H	0.799	0.799	0.799	14	Foreign
I	0.947	0.759	0.974	6	Public
J	0.938	0.707	0.983	8	Domestic
K	0.953	0.822	0.984	5	Domestic
L	0.933	0.534	0.979	9	Domestic
M	0.692	0.330	0.962	15	Domestic
N	0.942	0.898	0.968	7	Domestic
O	0.956	0.931	0.970	4	Domestic
All banks	**0.917**	**0.330**	**0.984**		

Source: Author's own calculations based on BoZ data

The overall picture emerging from the estimated cost efficiency indicators is that although Zambian banks are inefficient on average, most of them inched closer to the best practice frontier. This was shown by the average inefficiency gap of 8.0 percent for the whole sample period. The improvement in cost efficiency can be detected in Figure 4.1 below, which gives a graphical trend analysis of efficiency performance over the study period. Table 4.5 provides condensed annual summary statistics of efficiency indicators. To reiterate, the study period coincides with some positive developments in the economy, beginning particularly with 2002. These developments were marked by a rebound in economic growth and reduction in consumer inflation. Clearly, these changes strongly influenced the banks' cost performance response.

Figure 4.1: Evolution of mean cost efficiency



Source: Author's own computations based on BoZ data

Table 4-5: Annual summary statistics of cost efficiency

Year	Mean	Std Dev.	Min	Max
1998	0.847	0.181	0.330	0.966
1999	0.913	0.111	0.438	0.976
2000	0.909	0.115	0.506	0.977
2001	0.937	0.062	0.702	0.980
2002	0.920	0.062	0.739	0.980
2003	0.932	0.078	0.538	0.983
2004	0.897	0.092	0.668	0.982
2005	0.931	0.072	0.688	0.980
2006	0.957	0.023	0.893	0.984
1998-2006	0.917	0.099	0.330	0.984

In Figure 4.1, we observe an upward trend in cost efficiency, depicting some reduction in inefficiency over the sample period. The picture given by Table 4.5 for aggregate annual statistics of cost efficiency is somewhat blurred, mainly because of the variability in efficiency scores. Nonetheless, it reflects in broad terms the reduction in the degree of inefficiency among the sample banks. Based on the sample estimates, the mean efficiency index increased by about 13.0 percent between 1998 and 2006, an indication of narrowing cost inefficiency during the study period.

Although there were variations in individual banks' cost efficiency, the improvement in performance for all sample banks was remarkable and indicates that since the end of the crisis, the majority of banks have broadly moved closer to the best practice frontier although this shift is light. For star performers, survival strategies aimed at cost reduction and improved management practices partly explain the shift in cost efficiency. Therefore, from a policy perspective, the results show that liberalisation policies may have entered a new phase, injecting new lease of life in the Zambian banking sector. More importantly, the results represent the banks' strong response to incentives induced by the reforms towards stimulating managerial flexibility and functional reorganisation aimed at stemming costs and inappropriate mix of resource inputs.

4.6.4 Further analysis of cost efficiency and bank ownership structure

The analysis of results estimated from the inefficiency model suggested that foreign banks appeared less efficient than their domestic private and public sector counterparts. Figure 4.2 below puts this analysis into perspective by depicting cost efficiency by bank ownership structure. Theory predicts that foreign owned banks would under normal conditions perform better than domestic private and state banks mainly because of their international exposure and sound managerial practices.

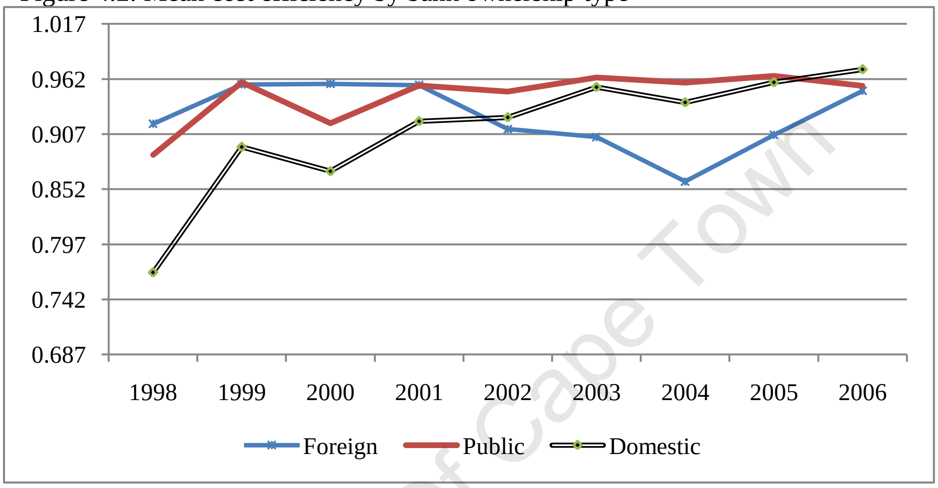
The analysis of Figure 4.2 corroborates econometric estimates which show that foreign and domestic private banks lagged behind public-owned banks in cost efficiency performance. Rather than depict improved cost performance, the relatively low level of inefficiency exhibited by public owned banks reflects more the cushioning effect of soft budget constraints in form of implicit guarantees and other incentives, including conversion of toxic assets related to state enterprises into long-dated government securities in order to shore up the public banks' balance sheets. For instance, the Zambia National Commercial Bank (ZNCB), the largest public bank, was losing revenue mainly due to non-performing loans. This required a provision of approximately K12.0 billion to keep it solvent.[33] According to official records filed by the Minister of Trade in Parliament (Government of the Republic of Zambia, 2007), ZNCB was also losing a further K2 billion (US$0.5 million) per month on interest due on deposits and in establishment and staff costs, with the large rural branch network contributing to this drain. Predictably, these losses were eroding the bank' capital base.

To rectify the situation and bring the bank back to equilibrium in order to facilitate the privatisation process, Government cleaned up its balance sheet and issued bonds valued at K250 billion (US$62.5 million) to cover the liabilities owed by the Zambia National Oil Company

[33] This converts into approximately US$3.0 million at the 2006 average exchange rate

(ZNOC) and a failed mining company, the Roan Mining Company of Zambia (RAMCOZ).[34] Inevitably, these actions strengthened the state bank's balance sheet and capital base and helped contain the costs related to bad loans, thereby improving its performance. However, these guarantees were of a short-term nature. To sustain its performance, long-term measures were necessary, including privatisation of the bank.

Figure 4.2: Mean cost efficiency by bank ownership type



Source: Author's own estimates based on BoZ data

Foreign banks experienced a downward trend in cost performance, which means inefficiency increased. The observed low efficiency levels by foreign banks may be due to less regulatory pressure enjoyed by these institutions. This enhances their reputation but also gives them greater latitude to maximize private benefits of control, including very generous staff compensation packages, resulting in high cost inefficiency performance. However, from 2004, cost efficiency of foreign banks recovered, as economic incentives changed following improvements in the economy. In contrast, efficiency of domestic private banks rose steadily throughout the study period, lifting up in 2002, coinciding with the improvement in broad macroeconomic performance. To this end, it can be surmised that domestic banks were quick to post gains from macroeconomic stability which characterised the period between 2002 and 2006.

[34] Prior to the conversion, 90 percent of loans were non-performing, indicating the severity of the liquidity position.

From 2004, the dispersion in cost efficiency between public and domestic private banks narrowed somewhat. Domestic private banks closed in on state banks' efficiency performance and exceeded the latter's cost efficiency by about 1.7 percentage points in 2006. However, the efficiency gap between foreign owned banks and domestic private banks was widening. Yao, et al. (2007) reported of similar efficiency behaviour among Chinese banks. They noted that the efficiency gap between Chinese state banks and domestic private banks ranged from 8 to 18 percent in favour of domestic private banks. The leap in cost efficiency of private banks was underpinned by the merger in 2004 between two of the small domestic banks. Before the merger, the average cost efficiency for the two banks was 70.0 percent. In the post merger period, cost efficiency for the bank formed by the two merged banks was 1.4 times higher than the average efficiency for the merged banks during the pre-merger period.

4.6.5 Empirical results of scale economies and technological progress

The main objective of this chapter centred on the analysis of cost efficiency and its determinants. However, efficiency performance may be related to banks' exploitation of scale economies or due to adoption of certain production technologies that help them reduce costs. Therefore, we devote the remainder of this chapter to the analysis of economies of scale and technological change in the Zambian banking industry. This discussion serves as a robustness check to the preceding analysis on characteristics of cost efficiency. Indices of global scale economies and technological change are reported in Table 4.6 below.

The results in Table 4.6 reveal existence of widespread economies of scale for all banks although they also show a downward trend over the estimation period. The estimated mean value of scale elasticity for all banks was 1.1, indicating that costs fell by 10 percent relative to the increase in the scale of production. Furthermore, the results exhibit a wide variation in the level of scale economies between small and large banks. Small banks were 12.6 percent more productive than large banks. The small banks subsample comprises mainly domestic private banks. Therefore, as earlier observed, the high indicator of scale economies may be partly explained by the relatively higher efficiency performance of this bank category.

Table 4-6: Economies of scale and technological change in Zambian banking

	1998	1999	2000	2001	2002	2003	2004	2005	2006	1998-2006
					Economies of Scale					
Small	1.278	1.250	1.228	1.206	1.185	1.167	1.152	1.132	1.110	1.190
Large	1.119	1.086	1.076	1.054	1.043	1.034	1.045	1.032	1.021	1.057
All banks	1.216	1.187	1.168	1.151	1.134	1.119	1.110	1.094	1.075	1.139
					Technological Progress					
Small	-0.014	-0.012	-0.009	-0.008	-0.004	-0.002	0.003	0.005	0.007	-0.004
Large	-0.010	-0.010	-0.006	-0.005	-0.002	0.001	0.006	0.008	0.009	-0.001
All banks	-0.013	-0.012	-0.008	-0.007	-0.004	-0.001	0.005	0.007	0.009	-0.003

Source: Author's own computations based on BoZ data

Parallel to the results on scale economies, estimates for technological progress indicate that for the sample period as whole, banks recorded slight cost gains due to technological progress. However, from 2004, the banking sector was characterised by technological regress averaging about 0.7 percent. Between 2004 and 2006, banks' costs increased but too small to offset scale efficiency posted in earlier years. For large banks, the deceleration in costs due to technological change was 0.8 percent, 3 basis points higher than that for small banks. This result underpins the fact that as technological progress took root, large banks adapted faster to these changes, although the existence of new technologies and diversification in services may not have spurred significant cost reductions over the sample period.

4.7 Summary of findings

In this chapter, we have analysed characteristics of cost efficiency for the Zambian banks using the stochastic frontier approach. We have also evaluated the importance of bank-specific and environmental (including regulatory) factors in explaining cost inefficiency of Zambian banks. The results from the analysis show that on average, Zambian banks were inefficient. This means that for the banking industry as a whole, mismanagement of resources remains an impediment to good performance. However, over the years, we observe some reduction in cost inefficiency, with domestic private banks displaying remarkable performance in this respect. In particular, this was evidenced in widening of the efficiency gap between domestic private banks and foreign owned banks.

Inefficiency in banking is influenced by a combination of factors. Most importantly, we have established that both internal and external variables affected cost inefficiency of Zambian commercial banks over the study period. Macroeconomic uncertainty, especially in the early years of the sample period played a greater influence in exacerbating the banks' cost inefficiency. Regulatory factors did not exacerbate cost inefficiency. It was found that compliance with cash reserves actually helped banks to improve cost efficiency. The same applies to regulatory capital. Well capitalised banks were also the most cost efficient. Internally, banks' credit risk exposure and reliance on Treasury securities dampened improvements in cost efficiency.

Evidence of widespread scale economies is also observed, although there was a declining trend over the sample period. As well, technological change characterised the banking sector, but cost savings were modest. Nonetheless, the analysis shows that large banks appeared to adapt faster to technological change than small banks, even though the difference in performance was less pronounced.

Chapter Five

Measuring the level of Competition in the Zambian Banking Industry

5.1 Introduction

The study of competition in banking markets spans at least four decades, yet this continues to be an attractive subject of empirical enquiry. From public policy perspective, banking competition represents a socially optimal target, since it stimulates reduction of intermediation costs and leads to delivery of high quality service by banks. Therefore, there are many reasons for studying competition in banking. Indeed the advantages of competition for an efficient and inclusive financial system are strong in that competitiveness fosters efficient management of banks thereby increasing firms' access to external financing (Beck, Demirgüç-Kunt, & Maksimovic, 2004). Related to this, competition in the banking sector can spur economic growth and lead to an improvement of social conditions.

A number of studies have shown the relationship between competition and growth. For instance, Pagano (1993) has argued that a competitive banking system leads to an improvement in economic growth and enhance consumers' social welfare. Cetorelli and Gamberra (2001) argue further that although concentrated banking systems offer growth opportunities for young firms, there is strong evidence of a general depressing effect on growth associated with a concentrated banking industry, which impacts all sectors and all firms indiscriminately. Claessens and Laeven (2005) reach similar conclusions arguing that the degree of competition is an important aspect of financial sector development.

In Zambia, the low level of access to banking services for all economic agents, but particularly households and small and medium sized firms is the subject of recurrent policy debate (de Luna Martinez, 2006; Oxford Policy Management, 2007; Ministry of Finance and National Planning, 2004). The main conclusion is that access to financial resources has been hampered by lack of competition in the Zambian banking sector. Therefore, lack of a competitive banking sector may be seen as the main cause of slow growth in the economy because firms have limited access to external finance to expand their businesses.

The seminal paper by Keeley (1990) stimulated the debate on the relationship between banking competition and financial stability. In his article Keeley (1990) argues that banks attach greater weight to charter value which then serves as a self-regulatory devise in that it prevents them from engaging in risk behaviour. Competition is therefore good for financial stability (Allen & Gale, 2004). However, the trade-off between competition and stability is a matter of intense debate. For instance, Hoggarth, et al. (1998) argue that competition in the banking sector leads to financial fragility because it induces banks to take on more risks. Using a dynamic theoretical framework, Bolt & Tieman (2004) show that banks faced with stringent capital adequacy requirements tend to set stricter acceptance criteria for loans and that stiff competition in the banking industry propagates riskier bank behavior thereby threatening stability of the financial sector.

On the other hand, evidence by Staikouras and Wood (2000) and Marcel, et al. (2001) shows that competition in banking does not endanger financial stability. Specifically, Marcel, et al. (2001) argue that where competition endangers instability in the financial sector, stability can be enhanced by prudential regulation and improved corporate governance. It may also be beneficial for a bank to hold equity capital in excess of regulatory threshold although this comes at a higher cost relative to deposits. According to Northcott (2004) the impact of banking competition on financial instability is a short-term phenomenon. In the long-term, firms and authorities must aspire to foster competition to ensure the future strength of the banking system. From a policy perspective, this can be achieved by designing regulatory measures that anticipate problems rather than those that react to them.

Arguing along similar lines, Beck (2008) shows that given the importance of a competitive banking sector to growth, regulatory and supervisory authorities should instead focus on devising an incentive compatible environment for banks rather than try to fine-tune market structure or the degree of competition. Therefore, competition must be placed at the centre of any public policy agenda since it has the mechanism to respond to the dynamic changes in economic conditions, especially those that affect delivery of financial services.

5.2 Problem statement and research objectives

A common expectation from financial liberalisation was that competition in the banking system would increase, thus allowing for reduced intermediation spreads and improved access to financial services by the people. In Zambia, Mwilwa (2007) observes that bank interest spreads have remained high years after the financial liberalisation reforms were implemented. Among the causes of high spreads, the author cites both bank-specific and the macroeconomic variables as being influential. High spreads may also be a reflection of collusive behaviour as well as managerial incompetence and inadequate technology (Gruben & McComb, 2003).

In Zambia, there is widespread perception that competition in banking is limited. This perception stems in part from the observed wide interest spreads and high prices of bank services in form of fees and commissions. These factors have tended to push banks' earnings nearer to the ones defined by oligopolistic market. Indeed, gauging by the high level of concentration in the Zambian banking industry, the concern for limited competition is warranted. Using these indicators, the IMF and World Bank (2002) also noted that these factors have undermined competitive behaviour in the banking sector.

Unlike the FSAP report by the IMF and World Bank, the FinMark Trust (2006) and study attributed lack of financial inclusiveness to less competition in the banking sector. de Luna Martinez (2006) uses a similar approach to make a case for low competition among Zambian banks. These studies posit that due to inadequate competition among banks, customers are subjected to a very high cost of banking services. Accordingly, this has shut out a large section of society from accessing financial services. Although there are other factors such as low incomes that contribute to poor accessibility to financial services, the cost of intermediation is certainly a major impediment.

Although there is a policy recognition that competition in the Zambian banking is low and creates a disincentive to improving efficiency performance of commercial banks, it is somewhat surprising that hitherto no empirical study has been conducted to understand the state of competition in the industry. The studies referred to above have drawn their conclusions by relying on methodologically flawed indicators which do not explain actual bank behaviour. Assessing competition using concentration ratios and related structural indicators leads to wrong inferences. This is partly because structural measures ignore the relationship between market contestability and revenue behaviour at bank level. To this end, a different form of analysis is required to capture the actual conduct of banks. Given the relative dominance of the banking industry in Zambia's financial sector, it is imperative that the market structure under which commercial banks operate is empirically assessed.

Therefore, the key research objective in this chapter is to assess the degree of competition in the Zambian commercial banking sector in the post reform period, particularly after the banking crisis of the mid-1990s. This is implemented using the Panzar-Rosse approach (Rosse & Panzar, 1977; Panzar & Rosse, 1982; Panzar & Rosse, 1987) anchored in the non-structural approaches, otherwise known as the New Empirical Industrial Organization (NEIO) models. The main appeal of the NEIO approaches, unlike their structural counterparts, is that they are a refinement of formal models of imperfectly competitive markets and endogenise market structure by taking into account the fact that bank performance may affect market structure.

Thus, of major importance in this exercise is the need to respond empirically to the continuing policy challenges and dilemmas presented by inconsistencies between expectations of policy reforms and observed evidence in the performance and conduct of the Zambian banks. Accordingly, the study is both timely and relevant to the present situation. The rest of the chapter is organised as follows. The next section reviews the relevant literature while section three discusses the methodology and data issues. In section four we present and discuss the empirical results. Section five gives a summary of the findings for the chapter.

5.3 Competition in banking – review of the literature

Two main approaches have dominated the study of competition in banking – the structural and non-structural approaches. Structural approaches draw significantly from traditional industrial organisation (IO) economics. The classic IO characterisation of banking competition has it that there is a causal link running from bank market structure to conduct and performance, measured by the banks' pricing behaviour and profits (Northcott, 2004). The leading methodology of structural models is the so-called Structure-Conduct-Performance (SCP) paradigm initially proposed by Bain (1951).

5.3.1 Structure conduct performance

The main prediction of the SCP paradigm is that a high level of concentration causes banks to charge high loan rates and low deposit rates, thereby reducing equilibrium quantities of funds available for credit (Besanko & Thakor, 1992). The majority of banking studies utilising the SCP analysis have been conducted in industrial countries, especially the United States and Canada. Among them, Diebold and Sharpe (1990); Hanan and Berger (1991); Hannan (1991a; 1991b); Neumark and Sharpe (1992) all found support of the SCP paradigm in banking. In particular, they predict that high loan interest rates, low deposit interest rates and occurrence of collusive profits are a function of market concentration. For a review of early studies on the SCP paradigm as applied to banking, see Gilbert (1984). Bikker and Haaf (2003) and Northcott (2004) give an updated review of literature.

Critics of the SCP paradigm have observed that the theory does not explain the sources of differential levels of market concentration such as presence of entry barriers. When these factors are controlled for, the evidence of the SCP paradigm disappears (Evanoff & Fortier, 1988). Further evidence strongly rejects the validity of the relationship between concentration on one hand, and interest rates on the other. For instance, Jackson (1992; 1997) has argued that there is no monotonicity between concentration and prices or interest rates. Instead, the evidence is U-shaped. Hanan (1997) also fails to find evidence of the impact of bank concentration on interest rates. In view of these criticisms, the SCP has lost much of its appeal in modern banking literature, yielding to alternative methods of analysing bank behaviour.

5.3.2 Efficiency structure hypothesis

Within the structural class of models, the efficiency structure hypothesis (ESH) due to Peltzman (1977) and Demsetz (1973) has challenged the predictions offered by the SCP. Proponents of the ESH submit that there is a reverse causality between competition and concentration, suggesting that higher profits may be a result of efficient operations of banks. Thus, banks may be exploiting greater X-efficiency, which in turn allows them to acquire market dominance.

The test for the ESH has also been extensive but with varying results. Rhoades (1992) investigated the degree of competition in the local mortgage loans market for the US. The results suggested that local market conditions were important in influencing banks' performance. Further, results indicate that the level of concentration indices significantly affected prices charged in the local mortgage markets, underpinning the relative importance of the SCP over the ESH explanation. Berger (1995) used a measure of X-efficiency in an attempt to distinguish between the SCP and ESH. The results were not conclusive. However, for emerging markets of Europe, Al-Obaidan (2008) found evidence confirming the efficiency hypothesis.

Frame and Kamerschen (1997) tested the possibility of superior X-efficiency in explaining high profits observed in the US legally protected banking markets. The authors refute that X-efficiency was the main reason US banks were highly profitable. This evidence was supported by Bernstein and Fuentes (2003) in their analysis of interest rate stickiness for the Chilean banking sector.

5.3.3 SCP and ESH literature for developing countries

Studies of bank behaviour in developing countries have always lagged behind research in developed countries and emerging markets. For countries in SSA, the paucity of literature is even more acute, both for the SCP and the ESH. To contextualise the discussion, we begin by looking at the recent evidence for Bangladesh, in which Samad (2008) tests the validity of the two competing hypotheses of the structural models. Using pooled analysis, he notes that the Bangladesh banking sector can be defined by the ESH. However, he suggests that there is need to further explore the effect of the banking structure in order to draw definitive policy conclusions.

In Africa and SSA in particular, Chirwa (2001; 2003) studied the market structure-profitability relationship for Malawian banks using cointegration analysis. For both studies, Chirwa argued that bank market structure matters for profitability in the Malawian banking system. A study by Okealaham (2004) for the banking sector of the Common Monetary Area (CMA) countries observed that concentration was an important factor in the banks' profitability. Nannyonjo (2002) tests the ESH for Ugandan banks and could not reject the hypothesis. However, Mugume (2007) refutes claims that the performance of the Ugandan banking sector is a function of efficiency hypothesis.

5.3.4 Non-structural approcahes to studies of banking competition

As can be observed from the review of empirical studies on the competing structural models of bank behaviour, the evidence is mixed. This lack of uniformity in results gave rise to further search for alternative approaches to better explain bank behaviour. Accordingly, non-structural approaches also known as the new empirical industrial organisation (NEIO) models have emerged and offer theoretically sound foundation of competition among rival firms (Panzar &

Rosse, 1987; Bresnahan, 1989; Bresnahan, 1982; Lau, 1982). These non-structural approaches are seen as reconciling the observed ambiguity between theory and firms' behaviour by appealing to the literature of oligopolistic markets in analysing industry competition (Cetorelli, 1999).

For the NIEO models, market concentration is not necessary to explain competition in the industry since competition can actually occur even in concentrated markets (Shaffer, 1994). Hence, as argued in the contestability literature (Baumol, 1982), the number of banks in the industry does not matter for competition. Rather, the driving force is the threat to entry defined by presence or absence of barriers, economic or legal.

5.3.5 A selection of studies for industrial countries

In the NEIO approaches, there are two leading methodologies, namely the Panzar-Rosse (PR) methodology and the Bresnahan-Lau (B-L) approach. Although both these models have been used to analyse the level of banking competition, a few studies have been conducted using the Bresnahan-Lau framework. Conversely, the PR approach has received wide application, both in industrial and nonindustrial countries. The main appealing feature of the PR approach is that it less data intensive and the knowledge of bank output and prices is not necessary. This has made it an attractive tool of investigating the state of competition in many countries. Besides, Bikker and Haaf (2003) observe that studies employing the Panzar-Rosse approach yield more consistent results than those relying on the structural methods of analysis.

Some of the early applications of the PR model to the banking sector in industrial countries were Nathan and Neave (1989) for Canadian banks, Vesala (1995) for Finnish banks, Molyneaux, et al. (1994) for a group of industrialised European countries. These studies all concluded that generally the banking industry is characterised by monopolistic competition. However, Nathan and Neave (1991) also found evidence of perfect competition in the Canadian banking industry for 1982 while Molyneaux, et al. (1994) noted that the Italian banking market could best be described by monopoly conduct.

Another authoritative study is that by Bikker and Haaf (2002) in which the authors apply the PR methodology to test the relationship between concentration and degree of competition for a number of industrial countries. The results are definitive and show that monopolistic competition was the dominant market structure for the banking sector of the sample countries. Another important finding was that the banking sector of the Netherlands was the most concentrated yet it was also the most competitive. Overall, Bikker and Haaf (2002) found a weak negative relationship between competition and concentration.

Casu and Girardone (2006) analysed the interplay among concentration, competition and efficiency for the single European market. They observed that banks were characterised by monopolistic competition and the degree of concentration did not seem to be necessary for banking competition. Matthews, et al. (2007) studied the competitive conditions of major UK banks and found that banks earned their revenue under conditions of monopolistic competition.

A distinguishing feature of the B-L approach is that it relies mainly on times series data, although panel estimations have been used in recent research. Notable studies in the B-L genre of literature include Shaffer (1993) who applied the Bresnahan-Lau model to the Canadian banking industry from 1965-1980. The results showed that banking behaviour was consistent with perfect competition during this period. Furthermore, Angelini and Cetorelli (2003) used a similar approach and derived a conjectural variation (CV) parameter to infer competitiveness in the Italian banking industry in the context of regulatory reforms. The authors observed that reforms induced improvement in the level of competition among regional Italian banks.

5.3.6 Bank competition in LDCs and emerging markets

As data have become readily available and economic reforms deepened, there has been a growing application of the NEIO methodologies in developing countries and emerging market economies, given that the link between traditional measures of competition and market structure disappear with deregulation of the banking system (Bruno, 2004). Here too the literature is replete with the Panzar-Rosse approach with a few exceptions.

One exception is Kubo (2006) who investigated the degree of competition for the Thai banking sector after the financial crisis based on the Bresnahan (1989) framework along the lines of Angelini and Cetorelli (2003). The results showed that banks had become less competitive after the crisis. Wong, et al. (2007) also derived a CV for Hong Kong during the period 1991-2005. The results indicated that despite rising concentration, banks in Hong Kong operated under competitive conditions in the loan market without any significant sign of collusive pricing.

A majority of studies in LDCs and countries in transition have popularised the use of the Panzar-Rosse approach. Many of these studies have been conducted in the context of financial liberalisation and other reform policies. For example, Mamatzakis, et al. (2005) employed the Panzar-Rosse methodology to a sample of South Eastern European (SEE) countries between 1998 and 2002. Their findings show that the banking systems in these countries were contestable, indicating absence of collusive behaviour. The level of competition had also improved over the years. A study by Gelos and Roldos (2004) for a group of European and Latin American emerging countries also found that contestability was high, despite rapid consolidation in the banking systems. Their Panzar-Rosse estimates of competition depicted presence of monopolistic competition in the banking industry of these countries.

For Greece, Hondroyiannis, et al. (1999), showed that banks operated in a monopolistically competitive market. Furthermore, gradualism in financial sector reforms and adoption of best banking practices in the context of the European Union directives were the main ingredients in propping up competitive conditions of the Greek banking system. These results were confirmed by Yildirim and Philippatos (2007) who argue that the banking markets of Central and Eastern European (CEE) countries could not be characterised by the bipolar cases of perfect competition or monopoly. Instead, banks earned their revenues under conditions of monopolistic competition. Moreover, the intensity of competition was higher in later years of the study period, indicating that banks may have been warming up to reforms. However, in an earlier study, Drakos and Konstantinou (2005) argued that banks in Greece and Turkey did not show significant differences before and after financial liberalisation although the finding for Turkey has been challenged. Specifically, Gunalp and Celik (2006) reported of improvements in the degree of competition among Turkish banks after liberalisation policies.

Whilst the research gap in many LDCs has been narrowing, the evidence for SSA countries has remained scant. For the few studies conducted, the broad observation corroborates findings of previous research elsewhere. Chen (2009) recently computed the competition index for a sample of middle-income SSA countries based on the Panzar-Rosse framework. For most the countries analysed, the results suggest that monopolistic competition best describes the level of competition in the banking sector although the degree of competition varies across countries. In another study, Saab and Vacher (2007) found that banking systems in the Economic and Monetary Community of Central Africa countries (CEMAC) are not integrated and are characterised by limited competition. This concern is shared by Buchs and Mathisen (2005) who noted that despite financial reforms, banks in Ghana continued to operate less competitively. However, Mugume (2007) and Hauner and Peiris (2008) found positive effects of reforms on the level of competition among Ugandan banks, especially in post-privatisation period. Clarke, et al. (2009) have observed that the privatisation of Uganda Commercial Bank has led to the improvement in the level of profitability for the acquiring foreign bank. However, the new bank still operated as a monopoly in remote areas due to low levels of competition.

Foreign bank presence in domestic banking markets has been found to shore up performance of financial institutions in host countries. The conventional argument for foreign bank presence in local banking markets is that they infuse new technologies, promote competition and provide managerial expertise. Foreign banks also increase local banks' resilience to shocks through peer pressure and learning by imitation. Also, the number of foreign banks by itself may exert a greater influence on the degree of competition even when their market share is large. This means that despite high levels of concentration, foreign banks may spur competition in the banking industry (Uiboupin, 2004; Claessens, Demirgüc-Kunt, & Huizinga, 2001).

A counter argument is that foreign banks tend to crowd out smaller domestic private sector banks in the deposits and loans markets. Because of their financial clout and reputation, foreign banks mobilise deposits from large multinational corporations, thereby crowding out domestic banks. In the loans market, large foreign banks also tend to lend to large firms without helping nurture the smaller domestic banks. This leads to a reduction of total credit available to small and medium-size enterprises (Stiglitz, 1994). In the short-term, foreign banks can also raise overhead costs for domestic banks and dampen their profitability as domestic banks seek to adopt new technologies and hire skilled personnel in order to compete with foreign banks (Claessens, Demirgüc-Kunt, & Huizinga, 2001).

With the exception of a few examples, see for instance Yuan (2006), and Yildirim and Philippatos (2007), many studies have not endeavoured to factor in the discriminatory effect of differences in bank sizes and ownership structure on banking competition. When such factors are taken into account, the evidence suggests that there are significant differences in the degree of competition between foreign and domestic banks and between large and small banks, respectively. In Zambia, large subsidiaries of foreign banks continue to dominate the banking industry in most segments of the banking market and even by size of assets (de Luna Martinez, 2006). This is because over the years, foreign banks have gained a substantial market share such that the entry of new fringe banks (both domestic and foreign) has not helped dampen the existing large banks' dominance.

5.3.7 Summary of the empirical literature

From the review of literature on banking concentration and competition we note that structural approaches suffer from theoretical deficiencies and therefore less equipped to offer plausible conclusions on the relationship between concentration and competition in banking markets. Shaffer (1994) cautioned that studies that infer competitiveness of the banking industry using structure or conduct are methodologically flawed, rendering their findings unsuitable as a basis for public policy. Indeed, as the analysis in the preceding section has demonstrated, banks can be competitive even when the market is highly concentrated.

Thus, the development of NEIO approaches to provide a theoretical basis to market behaviour has created further incentives for revisiting competition measures in the banking sector. From these approaches we observe that broadly, banks operate under conditions of monopolistic competition although a few cases of monopoly and perfect competition have been observed. Although it may be desirable to make comparisons across different banking markets, we should bear in mind that banks operate in uniquely different local economic conditions, and thus may respond differently to the opportunities and constraints presented by these conditions. Therefore, this aspect makes comparison in results problematic (Gunji, Miura, & Yuan, 2009). Each study must be evaluated on its own merit based on the economic environment in which commercial banks operate.

5.4 Methodology and estimation technique

5.4.1 Analytical framework of the Panzar-Rosse approach

The review of previous research has shown that the Panzar-Rosse methodology has theoretical and practical appeal as a measure of the state of competition in the banking sector. Therefore, in line with the literature, we adopt the Panzar-Rosse methodology in assessing the level of competition in the Zambian banking industry. Since this approach is widely used, the choice of the PR methodology enables us to compare our results with those obtained from previous studies, especially in developing countries.

The standard structure of the Panzar-Rosse methodology proceeds according to the analytical framework presented below and it is based on the structural revenue and cost relationship facing a particular firm, i. For details on formal derivation and accompanying assumptions, see Panzar and Rosse (1987). Let the revenue function be given by

$$R = R(y, \Omega) \tag{5.1}$$

where, y is output and Ω stands for other variables that shift the revenue function. Further, assume the cost function of the following expression

$$TC = C(y, w, \psi) \tag{5.2}$$

where TC denotes economic costs, w represents input prices, and ψ exogenous factors that affect banks' costs of production. These may also include factors internal to the banks' own operating environment. Profits are given by the difference between revenues and costs, implying that marginal revenue equals marginal costs. Therefore, the profit expression is denoted as

$$\pi = R(y, \Omega) - TC(y, w, \psi) \tag{5.3}$$

where π is per period level of profit and other variables are as defined earlier. Partially differentiating Equation (5.3) with respect to output, and equating the result to zero yields the profit maximization condition, that is, marginal revenue equals marginal cost as given by Equation (5.4)

$$\frac{\partial \pi}{\partial y} = \frac{\partial R(y,\Omega)}{\partial y} - \frac{\partial TC(y,w,\psi)}{\partial y} = 0 \quad \text{or} \quad \frac{\partial R(y,\Omega)}{\partial y} = \frac{\partial TC(y,w,\psi)}{\partial y} \tag{5.4}$$

where $\frac{\partial R(y,\Omega)}{\partial y}$ is marginal revenue (MR) and $\frac{\partial TC(y,w,\psi)}{\partial y}$ denotes marginal cost (MC).

Solving Equation (5.4) for y yields the profit maximising level of output, y^* defined as $y^* = y(\Omega, w, \psi)$. Substituting y^* into the revenue function yields the reduced form equation given by:

$$R^* = R\left[y^*(\Omega, w, \psi), \Omega\right] \equiv \tilde{R}(\Omega, w, \psi) \tag{5.5}$$

where $\tilde{R}$ denotes the reduced form revenue function and other variables are as defined before. It can be seen from Equation (5.5) that in long-run equilibrium, the firm's revenue is a function of factor prices and exogenous factors only. For the banking system, exogenous factors may include bank-specific, regulatory and institutional factors and macroeconomic variables. Clearly, neither the structural factors such as concentration ratios nor the number of firms is necessary to define the level of competition in the industry.

In the framework presented above, competition is measured by the extent to which a change in factor input prices is reflected in the equilibrium revenue earned by a bank. Panzar and Rosse (1987) show that the sum of revenue elasticities with respect to the firm's factor input prices indicates whether the market is perfectly competitive, monopolistically competitive or monopolistic (collusion). The sum of the factor price elasticities is given by the algebraic expression of the $H-$statistic in Equation (5.6) below:

$$H = \sum \frac{\partial \tilde{R}}{\partial w} * \frac{w}{\tilde{R}}. \tag{5.6}$$

Intuitively, the $H-$statistic rests solely on microeconomic theory, which exploits the proposition that pricing reactions to changes in input prices determine the market structure in which banks operate.

Thus, different market structures can be identified based on the sign of the $H-$statistic. Panzar and Rosse (1987) proved that if $H \leq 0$ the market is characterised by monopolistic conduct or conjectural variations short-term oligopoly. The implication is that an increase in input prices increases marginal cost, reducing the equilibrium output and hence revenue. Consequently, the sum of the revenue elasticities is negative. Under perfect competition or highly contestable market, an increase in factor prices leads to a proportionate increase in marginal cost and in equilibrium, this means marginal cost equals marginal revenue. Accordingly, there is no change in the equilibrium level of output indicating that the sum of elasticities is equal to one, that is, $H=1$. Monopolistic competition is defined by intermediate values of the $H-$statistic$\left(0<H<1\right)$. Intuitively, an increase in factor input prices raises a bank's marginal cost and hence revenue, but less than proportionately.

Table 5.1 below gives summary interpretations of the market structure and equilibrium conditions defined by the $H-$statistic.

Table 5-1: Interpretation of the Panzar-Rosse $H-$statistic

Value of $\boldsymbol{H}$ - statistic	Market Structure Characterisation
$H \leq 0$	Monopoly or conjectural variations short-term oligopoly. In this case each bank operates independently as under monopoly profit maximising conditions and the H-statistic is a decreasing function of the perceived demand elasticity.
$0<H<1$	Monopolistic competition characterised by free entry equilibrium excess capacity. The H-statistic is an increasing function of the perceived demand elasticity.
$H=1$	Perfect competition, or natural monopoly in a perfect contestable market, or sales maximising firm subject to break even constraint. It could imply free entry equilibrium with full (efficient) capacity utilisation.
Market equilibrium test	
$E=0$	Equilibrium
$E \leq 0$	Disequilibrium

Source: Molyneux, et al. (1994).

In the original formulation based on a single-product monopoly, Panzar and Rosse (1987) further showed that both the sign and magnitude of the $H-$statistic are important. That is, a higher $H-$statistic would indicate increasing competitiveness.

An important prerequisite of the PR methodology is that banks must be in long-run equilibrium for the above market structures to hold. For developed countries, the assumption of long-run equilibrium may not be difficult to sustain. In transition countries where banking sectors are still undergoing transformation, long-run equilibrium may be farfetched (Mktrtchyan, 2005; Northcott, 2004). However, given the internal logic of the model, it is best to think of equilibrium as a steady state, reflecting adjustment to shocks (Buchs & Mathisen, 2005).

For the Zambian banking system, the end of the crisis and subsequent restructuring of distressed banks culminated in relative stability of the banking industry. Moreover, over time there have been no major incentives for banks to continue to price their products and services below marginal cost in order to gain market share. These and related factors place the industry in what may be characterised as some sort of long-run equilibrium (implied steady state). Accordingly, Zambian banks may be seen as readily able to absorb the effect of exogenous shocks. Nevertheless, the assumption of long-run equilibrium must be empirically tested rather than imposed arbitrarily. This issue is addressed in the empirical section below.

Gelos and Roldos (2004) also argue that for the PR approach to be applied to banking, two other assumptions must be satisfied. First, banks should be modelled as single-product firms, using labour, capital and intermediated funds as inputs and secondly, input prices must be delinked from higher quality services because the opposite might imply higher revenues. When this is the case, the value of the $H-$statistic may be biased upwards.

5.4.2 Econometric estimation

For purposes of econometric estimation, the reduced form revenue equation is specified below. It is standard practice in the literature to use revenue, measured variously, as the dependent variable. The right hand side includes input prices, bank specific variables and macroeconomic or regulatory variables to control for market conditions. Thus, in line with previous research, see for example, Drakos and Konstantinou (2005); Buchs and Mathisen (2005), Hondroyiannis, et al. (1999), among others, the empirical PR model is given by Equation 5.7 below:

$$\begin{aligned}\ln\left(REV_{it}\right) = {} & \beta_0 + \beta_L \ln\left(w_{Lit}\right) + \beta_F \ln\left(w_{Fit}\right) + \beta_K \ln\left(w_{Kit}\right) + \beta_{RISK} \ln\left(RISK_{it}\right) \\ & + \beta_{OC} \ln\left(OPPCOST_{it}\right) + \beta_{CR} \ln\left(CAPRATIO_{it}\right) \\ & + \beta_{LR} \ln\left(INTERMED_{it}\right) + \beta_{BR} \ln\left(BRANCH_{it}\right) \\ & + \beta_{INFL} \ln\left(INFLATION_t\right) + \beta_{TBR} \ln\left(TBR_t\right) + \beta_{EXR} \ln\left(EXR_t\right) + \varepsilon\end{aligned} \qquad (5.7)$$

where subscripts i and t denote bank i at time t and REV is the ratio of revenue-to-total assets (both total and interest revenue as a proportion of assets are used), w_L, w_K and w_F denote respectively, unit labour costs (approximate wage rate), unit cost of capital and unit price of funds; $\ln$ is a natural logarithm operator, $OPPCOST$ is the opportunity cost of reserve requirements, $INTERMED$ is the intermediation ratio, $CAPRATIO$ is the capital-to-risk weighted assets ratio, $RISK$ is ratio of nonperforming loans to gross loans, $BRANCH$ is the number of bank branches per period, $INFLATION$ is the rate of inflation, TBR is the 91-day Treasury bill rate and EXR is the exchange rate defined as the amount of local currency units per unit of United States dollars, that is, ZMK/US\$, ε is a white noise disturbance term and T is time period measured in quarters. Finally, $\beta's$ are parameters to be estimated. From Equation (5.7), the $H-$statistic is given by the sum of the revenue elasticities with respect to input factor prices, that is, $H = \beta_L + \beta_K + \beta_F$.

All the independent variables have been carefully chosen and included in the revenue specification to reflect the banks' revenue response to regulatory constraints, internal management practices and macroeconomic conditions. For example, reserve requirements have the propensity to limit banks' independent behaviour and hence curtail competitiveness in the industry.[35] This is because they create compelling pressures that force banks to move in tandem yet independent behaviour is the hallmark of competition (Telser, 2007). Clearly, ignoring such factors may lead to misleading inferences about banks' actual behaviour.

---

[35] Reserve requirements are imposed and adjusted on occasion, mainly as a monetary policy tool than for prudential regulatory purposes. As a prudential regulatory requirement, they complement but not supersede equity and capital requirements.

5.4.3 Sample and data

All the 15 banks that were operating in Zambia during 1998-2006 form our sample of analysis. Newly established banks and merged banks during the study period have been included. However, the sample excludes failed banks due to unavailability of data. For the merged banks, the data are reported separately for each bank before the merger and subsequently, the merged bank enters as a single institution. Therefore, the estimations were conducted on an unbalanced panel.

Quaterly observations were used in the estimation process to increase degrees of freedom. The data were sourced from the banks' balance sheet and profit/loss statements as well as returns on liquidity ratios submitted to the central bank of Zambia. Macroeconomic data were gleaned from the Fortnightly Statistics, also published by the Bank of Zambia. Table 5.2 below gives a description of the variables used in the analysis. Note that only the newly introduced variables are presented in Table 5.2. Other variables used in the analysis are as defined in Table 4.1 above.

Table 5-2: Variables used in evaluating competition in the Zambian banking industry

Variable	Variable Name	Description and measurement	Mean	Median	Std Dev.	Source
$TREVASST$	Total revenue	Interest plus non-interest income expressed as a proportion of total assets	0.022	0.021	0.010	IS and BS
$INTRASST$	Interest revenue	Interest income, expressed as a proportion of total assets	0.013	0.013	0.006	IS and BS
$OITASS$	Other income	Non-interest income (fees, commissions, etc), expressed as a proportion of total assets	0.008	0.007	0.007	IS and BS
ROA	Return on assets	Total revenue before taxes expressed as a proportion of total assets	0.005	0.486	0.007	IS and BS
EXR	Nominal exchange rate	Local currency units per unit of US dollar (ZMK/US$)	3655.800	3806.965	1016.78	Fort. Stat.

Note: IS – Income statement, BS – Balance Sheet; Fort. Stat. – Fortnightly Statistics
Source: Bank of Zambia (BoZ)

5.4.4 Hypotheses and expected signs of coefficients

In this section we hypothesise about the direction of effect between revenue and its explanatory factors. Firstly, in line with previous research, we estimate the revenue equation using both total revenue-to-total assets ratio $(TREVASST)$ and ratio of interest revenue-to-assets $(INTRASST)$, taking the specification involving $TREVASST$ as the benchmark equation. Using total revenue to assess banks' behaviour ensures that a comprehensive measure of competitiveness for all banking services (traditional and non-traditional) is captured. This is consistent with the fact that for survival, banks have to aggressively contend with each other in all lines of business activities. Therefore, estimations based on interest revenue $(INTRASST)$ serve as robustness checks to the total revenue $(TREVASST)$ specification. Interest revenue comprises interest earned on loans, securities and net interbank placements. A regression for non-interest income has also estimated to reinforce the results from the two specifications.

All three input prices are expected to carry positive coefficients. Therefore, $\beta_L > 0; \beta_K > 0$ and $\beta_F > 0$ in all specifications. Since nonperforming loans are undesirable outputs, they act as a drain on revenue and bank performance. Accordingly, the variable $RISK$ is expected to carry a negative coefficient that is, $\beta_{RISK} < 0$. This implies that banks exposed to high credit risk would generate less revenue. A priori, the sign on opportunity costs of statutory reserves $(OPPCOST)$ is expected to be negative, $\beta_{OC} < 0$. The intuition is that a large amount of unremunerated reserves depict a high opportunity cost of funds. Accordingly, this means high interest forgone which in turn reduces the amount of revenue a bank would otherwise earn if such reserves were remunerated or invested in earning assets.

The effect of $CAPRATIO$ on revenue is ambiguous ex-ante, depending on whether high capital requirements lead to more or less revenue for the banks. To the extent that this variable has distortionary regulatory effects, bank revenue may decline, and therefore $\beta_{CR} < 0$. On the other hand, if such distortions are not binding, say, because banks engage is risk lending and hence earn a higher return, this could lead to more revenue. Accordingly, $\beta_{CR} > 0$. Banks that intermediate more funds into loans should earn a higher revenue. Therefore, $\beta_{LR} > 0$. We also expect that banks with more branches earn more revenue, due mainly to closer contact with customers. Accordingly, $\beta_{BR} > 0$.

High inflation is deterimental to banks' performance. Therefore, we expect a negative coefficient on $INFLATION$, that is, $\beta_{INFL} < 0$. In contrast, since banks have historically generated a significant amount of revenue from foreign exchange transactions, we hypothesise that $\beta_{EXR} > 0$.

However, to the extent that depreciation in the domestic currency signals macroeconomic uncertainty, this variable may have a negative impact on revenue, meaning that β_{EXR} would be negative. The 91-day Treasury bill yield rate (TBR) controls for the co-dependence between government domestic borrowing requirement and banks' performance. This variable is expected to carry a positive sign $(\beta_{TBR} > 0)$ indicating that investment in Treasury securities yields more revenue.

5.4.5 Test for long-run equilibrium

The empirical test for the long-run equilibrium condition is given by Equation (5.8) below. The dependent variable is the natural logarithm of the return on assets (ROA), while the explanatory variables remain the same as those given by Equation (5.7) above. According to Shaffer (1983) rates of return are equalised across banks and must not be correlated with input prices, that is, $E = 0$. Therefore, long-run equilibrium is tested using ROA on the premise that it is unrelated to input prices. Since ROA can potentially take on negative values, for convenience it is adjusted by a factor of one before taking logarithmic transformation. The specification for long-run equilibrium reads as follows:

$$\begin{aligned}\ln\left(ROA_{it}\right) = {} & \beta_0 + \beta_L \ln\left(w_{Lit}\right) + \beta_F \ln\left(w_{Fit}\right) + \beta_K \ln\left(w_{Kit}\right) + \beta_{RISK} \ln\left(RISK_{it}\right) \\ & + \beta_{OC} \ln\left(OPPCOST_{it}\right) + \beta_{CR} \ln\left(CAPRATIO_{it}\right) \\ & + \beta_{LR} \ln\left(INTERMED_{it}\right) + \beta_{BR} \ln\left(BRANCH_{it}\right) \\ & + \beta_{INFL} \ln\left(INFLATION_{t}\right) + \beta_{TBR} \ln\left(TBR_{t}\right) + \beta_{EXR} \ln\left(EXR_{t}\right) + \varepsilon\end{aligned} \tag{5.8}$$

where ROA is the banks' profit before tax expressed as percentage of total assets and $E = \beta_L + \beta_K + \beta_F = 0$ is the test for the long-run equilibrium condition against the alternative that $E < 0$.

5.5 Empirical results of the Panzar-Rosse approach

This section presents the empirical results of the Panzar-Rosse approach. Estimations were conducted using panel data analysis. The advantage of panel estimation over cross-section analysis is that it enables the researcher to combine time series and cross sectional data in order to account for unobservable individual bank-specific effects. Therefore it has great flexibility in modelling differences in behaviour across individual sample units (Green, 2003). There are two competing estimation techniques in panel data analysis - the fixed and random effects models. The choice between the two models is informed by the results of the Hausman specification test. In line with Hoechle (2007) and Green (2003) all regression equations were estimated with robust standard errors in order to correct for groupwise heteroscedasticity and cross-sectional correlation in panels.

5.5.1 Long-run equilibrium in banking

The validity of the Panzar-Rosse approach rests on the satisfaction of the assumption of long-run equilibrium. Results of this test are presented in Table 5.3 below. Based on the Wald statistic, we fail to reject the null hypothesis of long-run equilibrium in Zambian banking even at the conservative 10 percent level. The estimated $E-$statistic was found to be 0.073. The test result means that the Panzar-Rosse estimation procedure can be used to adequately evaluate the state of competition and related market structure in the Zambian commercial banking sector. In the remainder of this chapter we assess the degree of competition measured by the $H-$statistic.

Table 5-3: Long-Run equilibrium test estimates

Dependent variable: $\ln(ROA)$			
Independent Variables	Coefficient	t-statistic	p-value
Intercept	-1.564	-0.946	0.344
Input prices			
$\ln(w_L)$	-0.057	-0.601	0.549
$\ln(w_F)$	0.113	2.255	0.025**
$\ln(w_K)$	0.017	0.320	0.749
E - *Statistic*	*0.073*		
Bank-specific and regulatory variables			
ln(*OPPCOST*)	-0.147	-2.271	0.024**
ln(*CAPRATIO*)	0.102	1.128	0.260
ln(*RISK*)	-0.041	-1.315	0.189
ln(*INTERMED*)	-0.022	-0.329	0.742
ln(*BRANCH*)	0.373	2.295	0.007***
Macroeconomic variables			
ln*(INFL)*	-0.326	-2.962	0.003***
ln*(EXR)*	0.644	2.886	0.004***
ln*(TBR)*	0.265	2.295	0.022**
Diagnostics			
Hausman test (p-value)	0.070*		
σ_u	0.329		
σ_ε	0.421		
ρ	0.379		
R^2	0.033		
No. of Obs.	374		
F-statistic (p-value)	0.000***		
Null hypothesis: $E \leq 0$, Fail to reject $(p-value = 0.529)$			

Significance level: * p<0.10, ** p<0.05, *** p<0.01
Source: Author's own estimates based on BoZ data

5.5.2 Test of competition in banking: results of the total revenue equation

Table 5-4 below summarises random effects results of the PR regression with *TREVASST* as the dependent variable for the full sample. According to the results, the data fits the model reasonably well. Independent variables explain 64 percent of the banks' revenue ratio. As an alternative test for model adequacy, the $F-$statistic is also statistically significant at 1 percent.

Table 5-4: Competitiveness in the Zambian commercial banking industry

Dependent variable: $\ln(TREVASST)$			
Independent Variables	Coefficient	t-statistic	p-value
Intercept	-1.949	-2.628	0.009***
Input prices			
$\ln(w_L)$	0.295	6.146	0.000***
$\ln(w_F)$	0.272	10.417	0.000***
$\ln(w_K)$	0.118	3.455	0.001***
Bank-specific and regulatory factors			
ln(*OPPCOST*)	-0.064	-2.405	0.016**
ln(*CAPRATIO*)	0.003	0.109	0.913
ln(*RISK*)	-0.027	-1.694	0.090*
ln(*INTERMED*)	0.043	1.192	0.233
ln(*BRANCH*)	0.118	2.351	0.019**
Macroeconomics variables			
ln(*INFLATION*)	-0.132	-2.783	0.005***
ln(*EXR*)	0.176	2.229	0.026**
ln(*TBR*)	0.104	2.429	0.015**
$H-$statistic		0.685	
Null: H=0 (p-value) Monopoly		0.000***	
Null: H=1 (p-value) Perfect Competition		0.000***	
Market Structure	Monopolistic competition		
Diagnostics			
Hausman test (p-value):		0.954	
σ_u		0.193	
σ_ε		0.196	
ρ		0.492	
R^2		0.641	
No. of Obs.		379	
F-statistic (p-value)		0.000***	

Significance level: * p<0.10, ** p<0.05, *** p<0.01

Source: Author's own estimates based on BoZ data

The main attraction of the Panzar-Rosse methodology is the magnitude and significance of the $H-$statistic and the concomitant market structure depicted by this index. From Table 5-4 we observe that the estimated valued of the $H-$statistic of 0.69 obtained from the specification of total revenue-total assets ratio is significantly different from both zero and unit at 1 percent level.

This means that the null hypotheses of both monopoly and perfect competition are unambiguously rejected. Since this value lies between zero and one, the intuition is that over the study period, Zambian banks earned their total income under conditions of monopolistic competition. That is, for the banking system as a whole, individual banks had the ability to post higher revenue due to the exercise of some degree of market power.

The value of the estimated $H-$statistic is further buttressed by the significance of all three input factor prices and the plausibility of diagnostic statistics. The sample coefficients on all three input prices carry the expected positive sign and are significantly different from zero at the 1 percent level. Intuitively, this implies that factor prices are important for Zambian banks in the pricing of bank products and services. This result justifies the use of the intermediation approach in modelling bank behaviour by incorporating all cost elements – financial as well as real resource costs (Swank, 1996; Baltensperger, 1980).

The results also suggest that by magnitude, labour and funding unit costs contribute the most to the $H-$statistic. Moreover, most estimates of bank-specific and regulatory variables are evidently robust, corroborating prior expectations. The ratio of impaired loans to gross loans $(RISK)$ is significant at 10 percent and carries the expected negative sign. This means that credit risk does indeed impede banks' revenue performance although the impact is modest. Nonetheless, this credit risk effect possibly explains the reluctance of Zambian banks to expand their loan portfolio during the sample period. Banks with a wide reach in terms of branch density $(BRANCH)$ also generate more revenue. The significance of the estimated parameter for $BRANCH$ means that geographical diversification acts as an advantage for commercial banks in terms of revenue generation. Clearly, therefore, even though operating an extensive branch network attracts costs, proximity to customers is also beneficial to commercial banks.

Regulatory intensity related to reserve requirements constrained banks from optimising their revenue performance as shown by the negative and statistically significant coefficient on $OPPCOST$. Capital requirements denoted by $CAPRATIO$ did not impose a binding constraint on the revenue performance of commercial banks. In any case, the result shows that well capitalised banks also earned more revenue than their less capitalised counterparts although this effect is insignificant. Similarly, the banks' loan-deposit ratio $(INTERMED)$ was also found to be insignificant. This is consistent with the fact that the financial intermediation function of Zambian banks has been very low during study period.

From the results in Table 5-4, there is reasonable evidence to suggest that macroeconomic instability characterised by high inflation affected the banks' revenue performance of Zambian banks over the study period. The coefficient on *INFLATION* is negative and significant at 1 percent, implying that in high inflationary conditions, banks earn less revenue, as traditional sources of income get downgraded by inflationary pressures. However, bank revenues responded positively and significantly to the 91-day Treasury bill rate and the exchange rate. A unit increase in the *TBR* raised banks' total revenue by 0.1 percent and this effect was significant at the 5 percent level. The above finding underscores the fact that the banks' holdings of Treasury bills and Government bonds have contributed significantly to the growth in revenue over the years.

Commercial banks have remained the dominant players in the securities market, accounting for over 90 percent of total marketable securities holdings. Therefore, investment in securities has always been an important source of bank revenue, accounting for a third of total interest income. Furthermore, banks use the 91-day Treasury bill rate as a benchmark in setting their base lending rates. Hence, the 91-day Treasury bill rate is picking up all these effects. Similarly for the exchange rate, the results indicate that the depreciation in the domestic currency helped banks generate more revenue through realised exchange gains on the banks' foreign currency denominated assets, which in turn boosted the non-interest income component of total revenue.

Broadly, results of the effect of macroeconomic stance on bank revenue corroborate those reported in previous studies for developing countries where credit to government and macroeconomic instability influence the revenue performance of commercial banks. For example, Mamatzakis, et al. (2005), Hauner & Peiris (2008) and Yuan (2006) show that inflation was detrimental to performance of the banking sector but the impact was less significant. On the other hand, Buchs & Mathisen (2005) observed that for Ghanaian banks, the rate of inflation adversely affected banks' revenue generation while the Treasury bill rate acquired a different but robust interpretation, impacting bank revenue more positively.

The results presented above are consistent with those reported by Buchs & Mathisen (2005) who showed that both inflation and the Treasury bill rate affected revenue performance of Ghanaian banks in an important manner but in the opposite direction. However, studies by

Mamatzakis, et al. (2005), Hauner & Peiris (2008) and Yuan (2006) showed that although inflation deterred banks from increasing their revenue performance, its impact was less significant.

5.5.3 Economteric results of an interest revenue equation

This section reports the $H-$statistic estimated from the specification involving the ratio of interest revenue to total assets. Estimating the PR model using the ratio of interest revenue-to-assets ($INTRASST$) allows us to gain further insight into the degree of competition taking into account the fact that banks generate the bulk of their income from traditional banking activities. Therefore, this estimation addresses itself to the question of whether the degree of competition obtained with interest revenue differs substantially from that estimated using total revenue. If it is different, this may be attributed to the influence of other income (fees, commissions, etc.). Thus, as a further consistency check, we also estimated a regression equation for other income as a proportion of assets ($OITASS$). The results of both estimations are reported in Table 5-5 below.

Panel I in Table 5-5 reports the estimation results for interest revenue. From the results, the sum of the input factor elasticities was significantly larger than zero, rejecting the hypothesis of monopoly. Specifically, the $H-$statistic was calculated as 0.648. This figure was also found to be significantly less than unit ($p\text{-}value = 0.000$), which rules out the possibility of perfect competition in the traditional segment of the Zambian banking system. Fundamentally, these findings indicate the oligopolistic nature of the Zambian banking sector. Therefore, the results lend further credence to the earlier finding of monopolistic competition in the total revenue specification. Relative to the results obtained with the total income equation, the $H-$statistic calculated from an interest revenue equation is generally of similar magnitude.

A possible explanation for the minor difference in the degree of competition is because banks earn more of their total income in form of interest revenue (that is, from loans, securities and interbank fund placements). Accordingly, traditional sources of income roughly approximate overall activity in the banking sector, especially for a developing country like Zambia where the level of financial sophistication is limited.

Table 5-5: Competitiveness estimates with interest and non-interest income

	Dependent Variable					
	$\ln(INTRASST)$			$\ln(OITASS)$		
Variable	Coefficient	t-statistic	p-value	Coefficient	t-statistic	p-value
Intercept	-5.131	-8.051	0.000***	-0.668	-0.517	0.606
Input prices						
$\ln(w_L)$	0.252	5.391	0.000***	0.378	4.206	0.000***
$\ln(w_F)$	0.291	10.190	0.000***	0.121	2.605	0.010***
$\ln(w_K)$	0.105	2.609	0.009***	0.161	2.757	0.006***
Bank-specific and regulatory factors						
ln(*OPPCOST*)	-0.101	-3.257	0.001***	-0.117	-1.730	0.040**
ln(*CAPRATIO*)	-0.023	-0.728	0.467	0.048	0.140	0.394
ln(*RISK*)	-0.059	-4.068	0.000***	0.034	0.810	0.237
ln(*INTERMED*)	0.056	1.312	0.190	-0.095	-0.330	0.130
ln(*BRANCH)*	0.145	2.239	0.026**	0.006	0.043	0.966
Macroeconomic variables						
ln(*INFLATION*)	-0.194	-3.263	0.001***	-0.181	-2.002	0.046**
ln(*EXR*)	0.477	5.620	0.000***	-0.009	-0.061	0.952
ln(*TBR*)	0.202	4.163	0.000***	0.120	1.470	0.143
H-statistic	0.648			0.660		
H=0 (p-value)	0.000***			0.000***		
H=1 (p-value)	0.000***			0.006***		
Market Structure	Monopolistic Competition			Monopolistic Competition		
Diagnostics						
Hausman Test (p-value)	0.007***			0.000***		
σ_u	0.333			0.482		
σ_ε	0.212			0.366		
ρ	0.713			0.634		
R^2	0.441			0.338		
No. of Obs.	380			378		
F-statistic (p-value)	0.000***			0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent

$\ln(INTRASST)$ - Log of Total Interest Revenue/Total Assets

$\ln(OITASS)$ - Log of Other Income (Fees and Commissions, etc)/Total Assets

Source: Bank of Zambia and author's estimations

Generally, previous studies have produced mixed estimates of the $H-$statistic derived from the two revenue definitions. For example, Perera, et al. (2006) found that in South Asian banking systems the degree of competition inferred from the interest revenue regression was greater than that estimated from a total revenue specification. On the other hand, Buchs and Mathisen (2005) found that for Ghanaian banks, competition estimated from total revenue (*H=0.54*) was greater than that estimated from an interest revenue regression (*H=0.42*). Yildirim and Philippatos (2007) also report of varying degrees of competition for a number of Central and Eastern European (CEE) countries depending on whether the competitive index is derived from an interest revenue regression or a total revenue specification (see Yildirim and Philippatos, 2007:Table II, p.203). Chen (2009) presents a more consistent story of competitiveness based on the two revenue measures for a sample of middle-income SSA countries. Except for Cape Verde, all other countries considered showed that competition among banks was greater when assessed using interest revenue (see Table 5 in Chen, 2009). Hauner and Peiris (2008) reported relatively similar estimates for Ugandan banks.

As in the previous estimation with total income, other coefficients in the interest revenue equation are consistent with those for the total revenue regression. The unit factor prices carry expected signs with funding costs and unit labour costs dominating the influence in the $H-$statistic. The same is true for the three macroeconomic variables, namely inflation, Treasury bill rate and exchange rate, respectively. This means that bank interest income is subject to macroeconomic shocks and is strongly procyclical. For bank-specific and regulatory variables, the estimated coefficients have retained their signs and level of significance, except for the capital ratio ($CAPRATIO$), which has its sign reversed but remains insignificant. A noteworthy observation is the magnitude of the parameter estimate for $RISK$, which has more than doubled to 0.059 and is statistically significant from zero at 1 percent level in the interest revenue equation. This estimate shows that focussing on interest income brings out a pronounced influence of credit risk on bank revenue.

5.5.4 Are banks more aggressive in traditional income sources than fee income?

The finding that competitive conditions for total income and interest income are marginally different may suggest that fee-income and off-balance sheet activities are an insignificant component of the banks' business line. We test for this conjecture by estimating a separate regression for fee and commission based income, also expressed as a proportion of total assets ($OITASS$).

The Wald-statistic permits us to yet again reject the null hypotheses of monopoly and perfect competition at 1 percent. Most importantly, the estimate of the $H-$statistic of 0.66 suggests that this component of banking income is also defined by oligopolistic conduct (see Panel II of Table 5-5). The estimated competition index suggests that the state of competition from fees and other income is no less or more than it is with traditional measures of income. However, the lack of significance of most coefficients on control variables indicates that fee income responded less robustly to these factors, especially to internal bank-specific factors. These findings are difficult to understand given that most banks tend to charge fees and commissions tailored to the structure of their internal operations. The insignificance of the exchange rate is also surprising in view of the relative importance of foreign exchange gains in other sources of banks' income.

5.5.5 Ownership structure, size and degree of competition

In order to assess whether results from the full sample regression differ by from those defined by the behaviour of particular bank categories and size, the estimations were conducted on subsamples of domestic versus foreign banks and large versus small banks.[36] Foreign owned banks are all those foreign equity ownership in excess of 50 percent while domestic banks are defined as those whose majority shareholders are primarily of Zambian origin. At the end of 2006, there were eight foreign-owned banks and five domestic banks, including a state-owned bank. Regardless of the ownership structure, all banks are subjected to the same regulatory regime and are expected to comply with the regulatory requirements. Thus, it is anticipated that any difference in the level of competition between domestic and foreign owned banks would predominantly be due to intrinsic and unobservable characteristics defined by the form of ownership, for example, managerial ingenuity and expertise in generating revenue. For this reason and taking a cue from previous studies, we hypothesise that foreign owned banks would exhibit a higher level of competition than domestic banks.

[36] The subsamples for domestic and large banks include the Zambia national commercial bank (ZNCB), which until April 2007 was wholly owned by the Zambian government. However, as stated earlier, ZNCB often enjoyed soft budget support. In a sense, this made the playing field less even.

We also investigated whether or not the degree of competition among large banks differs from that of small banks, using total assets as a classification criterion for the size of banks. Thus, a bank is classified small or large depending on the proportion of its assets relative to the total industry assets.[37] We identified four large banks which accounted for over 80.0 percent of the industry's total assets. These banks had a mean asset ratio of 16.3 percent. One of these banks was state-owned, accounting for about a fifth of the industry assets. The remainder were foreign banks. Small banks had a mean asset ratio of only 2.2 percent but exhibited wide dispersion around the mean. Altogether there were 11 small banks more than half of which were domestic private banks. Small banks accounted for less than 19.0 percent of total industry assets. Obviously, there is some arbitrariness in such classification thresholds but this has the merit of keeping a balance of differently sized banks in the same asset category. As with large banks, there was an overlap in the small banks subsample, that is, the small banks had a mixture of foreign as well as domestic banks. A priori, large banks are expected to manifest stronger competition than their smaller counterparts which tend to operate in the fringe markets.

For purposes of brevity, results of the estimated $H-$statistic for each of the bank groups and asset class are summarised in Table 5.3 below (full regression results are presented in Appendix II). In order to compare the results of these panel regressions with findings from previous studies, the estimations were conducted only for total and interest income, again expressed as a proportion of total assets.

For the total revenue measure, the results indicate that foreign banks were more competitive than domestic banks. For foreign banks, the estimated value of the $H-$statistic was 0.68 and depicted monopolistic competition of the Zambian banking sector, given by the hypothesis test results. On the other hand, the competitive index for domestic banks was lower at 0.59. However, the statistical test also rejects presence of perfect competition and monopoly at 1 percent in favour of oligopolistic behaviour.

[37] This is the same classification used by the Bank of Zambia.

Table 5-6: H – statistic estimates by type of bank

		Local	Foreign	Small	Large
Log of Total Revenue/Total Assets (*TREVASST*)	H-statistic	0.591	0.684	0.668	0.533
Hypothesis Tests for Different Market Structures					
Monopoly $(H=0)$	*p-value*	0.000***	0.000***	0.000***	0.000***
Perfect Competition $(H=1)$	*p-value*	0.000*	0.000***	0.000***	0.002***
Log of Interest Revenue/Total Assets (*INTRASST*)	H-statistic	0.683	0.577	0.686	0.611
Hypothesis Tests for Different Market Structures					
Monopoly $(H=0)$	*p-value*	0.000***	0.000***	0.000***	0.000***
Perfect Competition $(H=1)$	*p-value*	0.004***	0.000***	0.000***	0.011***

Significance level: *** 1 percent, ** 5 percent and * 10 percent
Source: Bank of Zambia and author's estimations

The analysis of competitive conduct with an interest revenue equation produced a contrasting outcome. Although monopolistic competition still remains the valid market structure, the higher value of the H – statistic for domestic banks shows that these banks operate in a more competitive environment than foreign banks. This aggressive behaviour serves as a catalyst for competitive interest rate setting among the domestic banks, indicating that domestic banks set loan prices independent of subsidiaries of foreign banks but tend to get signals from their local counterparts.

On size and competition, the results in Table 5-6 reaffirm the fact that banks operated in a monopolistic competitive environment for both revenue measures and asset class. However, an important observation is that for both income measures, small banks had a slightly higher competitive edge over large banks. For the small banks subgroup, the H – statistic from the total revenue equation was 0.67 against 0.53 for large banks. With respect to the interest revenue equation, small banks had a competitive index of 0.69 compared with 0.61 for large banks.

The observed greater intensity of competition among small banks is inconsistent with prior expectations and findings of previous research. For example, Yildirim & Philippatos (2007), Bikker and Haaf (2002) and Belaisch (2003) argue that competition among large banks tends to be greater than that for small banks because large banks enjoy international exposure while small banks operate mainly in thin local markets where they enjoy market power. For the Zambian case, higher competitive pressures exhibited by the small banks subgroup underpin the fact that to survive, small banks have an established presence in urban centres where competition for customers is intense. Thus, because of relatively low reputation, small banks have devised several competitive mechanisms which enable them to retain their clients. One way is by offering competitive terms on loans and deposits to both large and small customers. On the other hand, a majority of large banks boast an extensive branch network. The alternative way of looking at this is that since the large banks subsample also includes the public sector bank, it may be the case that its inclusion in the regression analysis affected the results of this bank category.

The above results notwithstanding, the overall conclusion from the analysis of competition of different classes of banks corroborate the findings from the full bank sample estimations. This is to say that during the study period, Zambian banks operated in an environment defined by monopolistic competition.

5.5.6 Summary of findings and comparison with previous research

This chapter has addressed an important subject of competition and market structure in the commercial banking sector for a low-income country in the context of financial reforms but focussing largely on the post-crisis period. Using a not seen before detailed bank-level data set, the study draws on previous research anchored in the New Empirical Industrial Organisation literature. Specifically, we estimated the degree of competition based on the Panzar-Rosse $H-$statistic from reduced form revenue equations. The regressions were estimated for total income, interest revenue and other income, the latter two providing robustness check to the broad revenue measure.

The key finding from the analysis is that Zambian banks earned their income under conditions of monopolistic competition. This is provided by the significant estimates of the $H-$statistic which lie between zero and unit and significantly so for all three sets of revenue measures. Furthermore, the results showed a marginal difference between the degree of competition estimated from total income and interest revenue. The intuition being that interest revenue may be a close approximation of overall banking activity. This confirms the fact that traditional activities continue to be an important source of the banks' rent in developing countries. The results also justify the choice of the intermediation approach in modelling banks' conduct.

Regression results based on ownership structure and size of banks were broadly in line with full sample estimates of market structure. In all subgroup panel regressions, the hypotheses of perfect competition and monopoly were rejected at 1 percent significance level. However, the magnitude of the $H-$statistic depicted a mixture of the degree of competitiveness between different bank categories. For total revenue, subsidiaries of foreign banks appeared more competitive than domestic banks as shown by the value of the estimated $H-$statistic. However, this was not the case when competition was assessed using from an interest revenue specification. Foreign banks lagged behind domestic banks in the degree of competitive intensity indicating that the former group of banks may have been less aggressive in generating interest revenue. Possibly, this may indicate that domestic banks are defined by competitive pricing strategy aimed at capturing market share from foreign competitors.

The main observation from regression results based on the asset size of commercial banks was that small banks had a higher competitive index than large banks and this finding is invariant to the choice of revenue measure. The implication is that unlike in other countries, Zambian small banks exploit their urban presence to continuously compete in order for them to remain profitable. An alternative explanation is that by virtue of their size, large banks exert some degree of market power despite their wide reach in terms of presence and customer base. From a regulatory policy perspective, this evidence shows that authorities should indeed be concerned about the market dominance of a handful of large banks because this appears to have implications for competition.

The results obtained in this chapter are comparable to previous research findings, both for developed and developing countries. Specifically for LDCs, the measure of competition compares favourably with estimates of the $H-$statistic provided by Buchs & Mathisen (2005) for Ghanaian banks. They are also consistent with the estimates reported by Chen (2009) for middle-income SSA countries. Moreover, the degree of competition in the Zambian banking industry does not depart from the competitive indices reported for some advanced and emerging economies. For example, Drakos and Konstantinou (2005), Gelos and Roldos (2004), Bikker and Haaf (2002) have all concluded that banks operate in monopolistic competitive conditions.

However, estimates for Ugandan banks (Mugume, 2007; Hauner & Peiris, 2008) and South Asian banking markets as shown by Perera, et al. (2006) suggest relatively smaller values of the $H-$statistic than the evidence adduced in our study. This implies that Zambian banks exhibit a slightly higher degree of competitiveness than those in Uganda and South Asia. Suffice to say, such variations could be attributed to differences in economic conditions under which these banks operated. They could also be underpinned by progress made in microeconomic financial sector reforms and restructuring and how these impact on banking competition.

Chapter Six

Market Power and its Determinants in the Zambian Banking Sector

6.1 Introduction

The microeconomic theory of the banking firm offers different aspects on the conduct of banks and their pricing strategies. In particular, the existence of pure profits which stems from the exercise of market power depicts banks' long-run equilibrium configuration in an imperfect market situation. Traditionally, market power has been depicted by the Lerner Index (Lerner, 1934; Tirole, 1992) measured as a relative mark-up of price over marginal cost divided by price. Market power is especially prevalent in industries dominated by a few large firms, which serve as market leaders through collusive conduct. However, the evidence that dominance of large firms manifests itself in market leadership has been contested in the literature (Prince & Thurik, 1995). Market power may also arise from private firms' investment in technology and equipment aimed at improving the quality of products and services. In this regard, private ownership could be associated with higher price-cost margins, driven by better quality of services (Konings, Cayseele, & Warzynski, 2005).

There are a number of factors that influence banks' exercise of market power. They include among others, structural indicators such as concentration ratios, regulatory policy, cost efficiency and the macroeconomic environment in which banks operate. For the Zambian banking sector, sustained wide interest rate spreads, high levels of market concentration coupled with high profit indicators reinforce the view that Zambian banks exercise market power in pricing bank products and services.

Studies have shown that under certain conditions, there is a close link between the H-statistic of the Panzar-Rosse methodology and the Lerner Index (Shaffer, 1983; Bikker & Haaf, 2002). The reasoning is that when banks are faced with a constant elasticity of demand for banking products and assuming there is no mismeasurement of variables, there is a symmetrical relationship between the H-statistic and the Lerner Index. Hence, as banks become more competitive (higher H-statistic), the narrower the relative price-cost mark-up (Lerner Index).

6.2 Study objectives and hypothesis

There are two main objectives in this chapter of the thesis. The first objective is to provide evidence of exercise of market power by Zambian banks and analyse its evolution over the sample period. The second objective is to investigate the factors that explain banks' market power. This is accomplished by relating the estimated bank-specific Lerner indices on market structure, bank-specific and environmental factors (regulatory and macroeconomic variables). This is of particular importance to regulatory authorities that rely on defective measures of market power. To the extent that a deregulated environment helped banks retain market power despite policies of open entry into the sector, we could witness higher price-cost margins in the Zambian banking industry. Since the restructuring programme was aimed at dealing with the weak small banks, this process could have reinforced the large (foreign) banks' dominant position. In particular, it might have acted as an incentive for subsidiaries of foreign banks to maintain a grip on their market share.

To the best of our knowledge, our study is the first of its kind to be conducted for the Zambian banking industry. A recent attempt to model bank behaviour by Mwilwa (2007) used an aggregate approach to the analysis of interest spreads. However, unlike Mwilwa (2007), our study utilises a unique rich bank level panel data set to analyse pricing behaviour in the Zambian commercial banking sector. In this regard, the availability of longitudinal data allows us to account for unobservable individual bank differences, which, although likely to be important, were not considered in Mwilwa's (2007) study. Therefore, the main contribution of this study to overall literature and to the Zambian banking sector in particular lies in the estimation of the Lerner Index using the output price and marginal cost estimated from a cost function. Hence, the study narrows the research gap on the measurement of banks' market power in developing countries and particularly in SSA. For SSA, there is especially dearth of systematic research on the behaviour of banks that is comparable and consistent with the research in developed countries and emerging markets. The exception is Aboagye, et al. (2008) for the Ghanaian banking sector, a study which we use as a benchmark for comparing our findings. These issues are taken up in detail in the remainder of this chapter.

6.3 Overview of the literature on market power

Perfect competition and monopolistic conduct do not adequately provide practical explanations of banking behaviour. This has resulted in alternative models of analysing banks' conduct. One such perspective relates to the banks' exercise of market power defined by the Lerner Index. Coccorese

(2009) has argued that the Lerner Index is a true reflection of the banks' degree of market power by depicting the behavioural departure from monopoly and perfect competition.

6.3.1 Theoretical litetaure of market power

Models of oligopoly behaviour have become increasingly popular in analysing bank conduct, including market power. These imperfect competition models offer robust improvements to the traditional measures of banking conduct such as the SCP and they make the analysis more appealing (Toolsema, 2004; Freixas & Rochet, 1997). The starting point of these models is that banks operate under conditions of imperfect competition defined by oligopolistic behaviour. A theoretical framework for analysing bank profit margins has its roots in the seminal works of Klein (1971) and Monti (1972) and the models of banking discussed earlier.

6.3.2 Empirical literature on market power in banking

Previous studies of bank behaviour relied on aggregate indicators of performance. However, in recent years, firm level studies have emerged thereby remedying the inadequacies inherent in aggregate analysis. The growth in firm-level panel studies has been spurred by the availability of individual firm data and the discontent about the failure of structural models to explain bank behaviour. By analysing firm level mark-ups, one can readily satisfy the assumptions of possible price-taking behaviour among individual firms as opposed to previous studies which violated this assumption (Hanan & Liang, 1993). The use of bank-level data also provides an opportunity to examine the evolution of market power over time and across firms. An assessment of the efficacy of regulatory and other reforms in banking has also informed a larger body of studies on market power.

In view of the above, Angelini and Cetorelli (2003) analysed the behaviour of Italian regional banks using a Lerner Index estimated from a conjectural variations model. The authors found that deregulation fostered a reduction in price-cost margins, consistent with theoretical predictions. Specifically, the authors showed that initially, competitive conditions were relatively unchanged, but improved substantially after the reforms, as evidenced by a reduction in the Lerner Index. However, the study failed to find support of the effect of the consolidation process on market power.

Fernandez de Guevara, et al. (2005) estimated the Lerner Index but could not find significant improvement in competitive conditions within the banking markets of the European Union (EU) member countries. To assess the determinants of market power, the authors included among other factors, a measure of concentration in the deposits market, which was found to be insignificant. However, the authors found that efficiency, default risk and bank size were strong explanatory factors of price-cost margins in EU banking markets. More recently, Fernandez de Guevara and Maudos (2007) estimated the Lerner index for Spanish banks following the approach of Fernandez de Guevara, et al. (2005). Their conclusion was that market power increased in among Spanish banks, driven largely by bank size, efficiency and specialisation. However, bank concentration was found to be an insignificant variable; the same observation made by Fernandez de Guevara, et al. (2005) for the banking sector in the EU.

In a study of market power in Swedish banking, Sjöberg (2006) estimated a conduct parameter based on the Bresnahan (1982; 1989) oligopoly model. The estimated Lerner Index indicates that the degree of competition was especially high (lower market power) among the large banks in Sweden despite high concentration in the sector. Based on the same framework, Fischer and Hempell (2006) showed that the Lerner Index for German banks depicted increasing competitive pressures with regional structural and economic variables playing an important role in sustaining the banks' exercise of market power. Demand factors were also found to have a strong economic effect on market power but the level of concentration was insignificant. Pruteanu-Podpiera, et al. (2008) analysed the degree of market power among Czech banks. The study did not find a clear cut evolution of the Lerner Index in the credit market, despite a high degree of concentration in the Czech banking industry.

Risk measures may affect bank behaviour in an important way. Oliver, et al. (2006) test this hypothesis by incorporating a measure of default risk in their analysis of market power in the Spanish credit market. Using bank-level data, the authors used a risk adjusted market interest rate to capture what they term a true measure of marginal opportunity cost of funds. They then derived a Lerner Index defined as a relative mark-up of loan rate over the risk adjusted market rate. Their results show that market power differs markedly across different loan products and that adjusting for risk premium significantly lowers the estimates of the Lerner Index.

Although this approach has appeal and produces some interesting estimates of market power for the Spanish credit markets, the methodology poses a challenge for developing countries where data on actual values of default risk are virtually non-existent. Consequently, many studies resort to using ex-post risk measures such as non-performing loans and loan loss provisions as approximate measures of bank credit risk. The degree of bias inherent in using such measures as determinants of market power is an empirical issue although there appears to be consensus that credit risk augments banks' exercise of marker power.

The literature of bank market power and its determinants in developing countries is sparse and mainly confined to emerging and transition economies. For sub-Saharan Africa, (SSA) the evidence is especially scanty. The majority of the studies infer market power using interest rate spreads, implying that these are good measures of market power. The problem with this approach is that market power may be endogenous. For instance, interest rate spreads could be affected by banks' exercise of market power while high margins may themselves by a consequence of market power, as demonstrated by Moore and Craigwell (2002), Chirwa and Mlachila (2004) and Vera, et al. (2007), among others. Other studies such as Mwilwa (2007) and Ngugi (2001) have taken an aggregate time series approach in analysing the determinants of bank interest spreads without exploiting the benefits of individual banks' interest rate setting behaviour tenable only through panel estimation.

As a remedy to the above shortcomings, Fahrer and Rohling (1981) argue it is important to apply approaches that take into account the direct behaviour of commercial banks in estimating market power. This is the basis of the growth in the NEIO based models for estimating market power. The NEIO approaches recognise the need to endogenise market structure in the banking industry and test the exercise of market power without relying on structural measures such as concentration ratios and number of firms (Delis, Staikouras, & Varlagas, 2008). Indeed competition can exist even in a duopoly market while monopolistic conduct is also possible even in markets with a large number of players. Ausunbel (1991) has shown that the U.S. credit card market with a large number of players was characterised by abnormal returns and asymmetric power reminiscent of monopolistic behaviour.

Only a handful of studies for SSA have used NEIO approaches in measuring market power more directly. Specifically related to our study is a recent paper by Aboagye, et al. (2008) for Ghanaian banks. The authors applied a methodology employed along the lines of Fernandez de Guevara & Maudos (2007) and Fernandez de Guevara, et al. (2005) to a panel of Ghanaian banks. Their observation is that Ghanaian banks possess market power stemming from the size, efficiency and the macroeconomic environment in which they operate. The other study by Okealaham (2007) took a different approach to the assessment of market power in the banking sector. The author argues that banks' exercise of market power is not a reflection of market imperfections but a consequence of cost economies. This implies that commercial banks enjoying scale economies may exercise greater market power thereby dwarfing the effects of structural indicators such as concentration ratios. Therefore, the author cautions that consideration must be given to cost economies and profitability in banking when drawing conclusions about market power effects.

Other studies for developing countries include Solis and Maudos (2008), who estimate and offer evidence on the social cost of market power using data from the Mexican banking sector. The authors found that banks exercised substantial market power in setting loan interest rates with the consequence that cost efficiency was significantly undermined. The estimate of the Lerner Index derived from interest rates showed that in 2005, social cost of market power was 0.15 percent of GDP. However, no evidence of the 'quiet life' hypothesis was found for the deposits market.

Tovar, et al. (2007) analysed the interplay between risk, concentration and market power in the Colombian banking industry after the regulatory reforms. Their results show that subsequent to the reforms, Colombian banks exhibited high systemic risk, which served as a collusive device and the cost of this risk was borne by bank customers through high mark-ups. However, the authors found that market concentration was not a significant collusive factor; rather, its effect was only robust after controlling for systemic risk.

6.4 Methodological framework

In this section we present a model for analysing bank market power, drawing largely from previous studies. A condensed version of the Klein-Monti framework is first presented before outlining the empirical structure of the cost function to be estimated.

6.4.1 The analytical structure of the model

The basis of the empirical estimation is the influential Klein-Monti theoretical oligopolistic model of banking first introduced in chapter three and reproduced here as a matter of pedagogy. Some of the leading empirical studies in this genre of models include Corvoisier & Gropp (2002); Gropp, et al. (2007) and Fernandez de Guevara, et al. (2005). Another variant of imperfect competition models stems from the conjectural variations approach following Iwata (1974), Appelbaum (1982), Bresnahan (1982; 1989) and Lau (1982). Applications of this group of models have included Shaffer (1993; 2001), Angelini and Cetorelli (2003) and Kubo (2006), among others.

In the ensuing discussion, we replicate the Klein-Monti model as the framework for the analysis of market power. For simplicity, we assume a case of duopoly in line with Dvořák (2005). Let the cost function be denoted by $TC_i\left(D,L\right)=\gamma_D D_i+\gamma_L L_i$, $i=1,2$ where TC, as before, represents total operating costs, D and L denote total deposits and loans while γ_D and γ_L are marginal costs of producing them, respectively. The loan rate $\left(r_L\right)$, deposit rate $\left(r_D\right)$, and the money market rate $\left(r\right)$, are exogenously determined since banks are assumed to compete in quantity. Therefore, each bank faces a downward sloping demand curve for loans and an upward supply curve for deposits, given by $L\left(r_L\right)$ and $D\left(r_D\right)$, respectively. The loan and deposit rates are inverse functions of the demand for loans and supply of deposits, as depicted by $r_L\left(L\right)$ and $r_D\left(D\right)$, respectively. Finally, the cash statutory reserves denoted by R, are given by $R=\left(1-\alpha\right)D-L$ where α is a fraction of deposits held as cash reserves at the central bank. Taking the amount of loans and deposits chosen by other banks as given, each bank maximises its profit according to the following specification

$$\pi_i=\left\{\left[\left(r_L\left(L_1+L_2\right)-r\right)\right]L_i+\left[r\left(1-\alpha\right)-r_D\left(D_1+D_2\right)\right]D_i-\left(\gamma_D D_i+\gamma_L L_i\right)\right\} \tag{6.1}$$

where, π_i is bank i's profits, L_1,L_2 are the amount of loans granted by bank 1 and 2 whereas D_1,D_2 are the amounts of deposits received by bank 1 and 2, respectively. Other variables are as defined above. Equation (6.1) shows that a bank's profit can be expressed as the difference between intermediation margins and operating expenses. The first order conditions for loans and deposits are derived as follows

$$\frac{\partial \pi_i}{\partial L_i} = \left(r_L^{'}\left(L^*\right)\frac{L^*}{2} + r_L\left(L^*\right) - r \right) - \gamma_L = 0$$

$$\frac{\partial \pi_i}{\partial D_i} = \left(r\left(1-\alpha\right) - r_D\left(D^*\right)\right) - r_D^{'}\left(D^*\right)\frac{D^*}{2} - \gamma_D = 0 \qquad (6.2)$$

Since L^* and D^* are solutions to the bank's objective function, they denote equilibrium amounts of loans and deposits for the banking sector. Rewriting the first order condition for loans in elasticity form yields the now familiar Lerner Index $\left(LI\right)$ given by equation (6.3) below

$$LI = \frac{r_L^* - \left(r + \gamma_L\right)}{r_L^*} = \frac{1}{2\varepsilon_L\left(r_L^*\right)} \qquad (6.3)$$

where ε_L denotes the elasticity of demand for loans and LI is the Lerner Index defined above. The only difference between this setup and the case of N-firms presented in chapter three is that for the duopoly case, the demand elasticities are scaled by a factor of two.

6.4.2 Empirical model and estimation strategy

The estimation of the Lerner Index requires knowledge of the price of banking output such as realised loan interest rates. However, for the banking industry in general and the Zambian banking sector in particular, the data on loan interest rates are not readily available. Since the flow of banking services is proportional to its stock of assets, the price of bank output can be depicted by the ratio of total revenue (interest and non-interest income) to total assets. Averaging across all banks yields an average market price. Using total revenue as a percentage of total assets to construct output price ensures that on-balance sheet outputs and off-balance asset items are all captured to avoid understating bank production (Carbó Valverde, Humphrey, & Rodriguez, 2003; Jagtiani & Khanthavit, 1996). It is therefore consistent with overall bank behaviour.

The profit and loss statements do not also report marginal costs related to the production of any of the individual asset items for Zambian banks. Therefore, an approximate measure of marginal cost has to be estimated. Given that marginal cost is not directly observable, it has to be estimated from an observable translog cost function. In industrial countries and other developed emerging markets where interbank trading accounts for a large volume of sources of funds for commercial banks, the interbank interest rate has been used as proxy for marginal cost of producing loans. However, in less developed countries rudimentary financial markets, the main source of banks' funds is through deposit mobilisation rather than the interbank market. Specifically in Zambia, the amount of funds raised through the overnight interbank market represents only 1.8% of total liabilities compared while deposits account for 78.2%.[38] While recognising that banks are multioutput firms, we aggregate all bank outputs, namely loans, securities, and other assets into an aggregate measure of bank product (total assets) which enables us to construct a single measure of marginal cost for overall production activity. If costs were reported for individual bank products and markets, we would then calculate output-specific marginal costs as Berg and Kim (1998) did for retail and corporate submarkets. With these caveats in mind, we replicate the empirical translog cost function of chapter four, yielding Equation (6.4) below:

$$
\begin{aligned}
\ln TC_{it} &= \beta_0 + \beta_y \ln Y_{it} + \frac{1}{2}\beta_{yy}\left(\ln Y_{it}\right)^2 + \beta_1 \ln w_{Lit} + \beta_2 \ln w_{Fit} + \beta_3 \ln w_{Kit} \\
&\quad + \beta_{11}\left(\ln w_{Lit}\right)^2 + \beta_{22}\left(\ln w_{Fit}\right)^2 + \beta_{33}\left(\ln w_{Kit}\right)^2 + \beta_{12} \ln w_{Lit} \ln w_{Fit} \\
&\quad + \beta_{13} \ln w_{Lit} \ln w_{Kit} + \beta_{23} \ln w_{Fit} \ln w_{Kit} + \sum_{j=L,F,K} \delta_{yj} \ln Y_{it} \ln w_{jit} \\
&\quad + \theta_t t + \theta_{tt} t^2 + \sum_{j=L,F,K} \theta_{tj} t \ln w_{jit} + \theta_{ty} t \ln Y_{it} + \psi_1 \ln\left(BRANCH_{it}\right) \\
&\quad + \psi_2 \ln\left(RISK_{it}\right) + \psi_3 \ln\left(INTERMED_{it}\right) + \zeta, \\
& i = 1,\ldots,N;\ t = 1,\ldots,T
\end{aligned}
\tag{6.4}
$$

where, as before TC_{it} denotes total operating costs, w_{jit} represents factor input prices, Y_{it} is total output, j,m denote actual inputs, ζ is an error term. Other variables are as defined in previous

[38] Although the interbank rate may be reflective of funding cost trends in industrial countries, in Zambia it does not represent the true marginal cost of bank production. Therefore, this study has opted to use a broad measure of marginal cost which takes into account all bank expenses, including cost of interbank funds, operating costs, and more importantly the cost of deposits and other borrowed funds. Accordingly, there is no loss of generality by setting aside the interbank rate as a measure of marginal costs of bank production.

chapters. In line with the literature and following the discussion in chapter four, we imposed symmetry and homogeneity conditions on the cost function above. Homogeneity condition is imposed by dividing total costs, the unit price of labour and the price of capital by the unit price of funds.

The cost function depicted by Equation (6.4) can be estimated directly using ordinary least squares. However, there are efficiency gains when the cost function is estimated jointly with input cost share equations. The inclusion of the cost share equations in the estimation procedure has an advantage of creating more degrees of freedom without adding any unrestricted regression coefficients. It also yields more asymptotically efficient parameter estimates than would otherwise be if the cost function is estimated on its own. The relationship between the cost function and the input share equation is implied by duality theory and can be derived using Shepherd's Lemma. Input share equations for labour, funds or capital can then be obtained by partially differentiating the cost function with respect to w_L, w_F or w_K. Input share equations for the three inputs are given below:

$$\begin{aligned}
S_L &= \frac{\partial \ln TC_{it}}{\partial \ln w_{Lit}} = \frac{w_{Lit} X_{Lit}}{TC_{it}} = \beta_1 + \beta_{11} \ln w_{Lit} + \beta_{12} \ln w_{Fit} + \beta_{13} \ln w_{Kit} + \delta_{yL} \ln Y_{it} + \theta_{tL} t + \omega_j, \\
S_K &= \frac{\partial \ln TC_{it}}{\partial \ln w_{Kit}} = \frac{w_{Kit} X_{Kit}}{TC_{it}} = \beta_1 + \beta_{22} \ln w_{Kit} + \beta_{21} \ln w_{Lit} + \beta_{23} \ln w_{Fit} + \delta_{yK} \ln Y_{it} + \theta_{tK} t + \omega_j, \\
S_F &= \frac{\partial \ln TC_{it}}{\partial \ln w_{Fit}} = \frac{w_{Fit} X_{Fit}}{TC_{it}} = \beta_1 + \beta_{33} \ln w_{Fit} + \beta_{31} \ln w_{Lit} + \beta_{32} \ln w_{Kit} + \delta_{yF} \ln Y_{it} + \theta_{tF} t + \omega_j,
\end{aligned} \qquad (6.5)$$

such that $S_L + S_K + S_F = 1$ and ω_j is a random error term for the jth input $\left(j = L, K, F\right)$.

Since the input shares sum to unity, one of the factor share equations is dropped to obtain a non-singular covariance matrix. Therefore, only $J-1$ share equations are estimated jointly with the cost function. The resulting parameter estimates are asymptotically equivalent to those obtained by the maximum likelihood approach and are invariant to the factor share equation dropped during estimation. The effects of variations across banks are reflected in the stochastic disturbance term, ω_j related to each share equation. These disturbances are assumed to be identically and independently distributed with mean zero and constant variance, that is, $\omega \sim N\left(0, \sigma_\omega\right)$. They are also orthogonal to the regressors. Furthermore, allowing disturbances to be correlated across equations enables us to apply Zellner's (1962) two-step iterated seemingly unrelated regression estimation (ISURE) procedure to the system of equations comprising the cost function and $J-1$ share equations. Since the cost function is normalised by the input price of funds, we have two share equations for labour (S_L) and capital (S_K), respectively.

Partially differentiating Equation (6.4) with respect to $\ln Y_{it}$ yields a measure of marginal cost according to Equation (6.6)

$$MC_{it} = \frac{\partial \ln TC_{it}}{\partial \ln Y_{it}} = \frac{TC_{it}}{Y_{it}}\left(\beta_y + \beta_{yy}\ln Y_{it} + \sum_{j=L,K}\delta_{yj}\ln w_{jit} + \theta_{ty}t\right) \tag{6.6}$$

where MC_{it} is the bank level marginal cost. Consistent with standard literature, see for instance, Angelini and Cetorelli (2003), Fernandez de Guevara, et al. (2005) and Aboagye, et al. (2008), the bank specific Lerner Index measure of market power is given by

$$LI_{it} = \frac{p_{it} - mc_{it}}{p_{it}} \tag{6.7}$$

where, p_i denotes output price for each bank, proxied by the ratio of total revenue to total assets for each period. Equation (6.8) below depicts the average Lerner Index for the banking industry

$$LI_t = \frac{P_t - MC_t}{P_t} \tag{6.8}$$

where P captures the market price for the whole banking industry, calculated by averaging p_i across all the banks for each period. The Lerner Index measure of market power may also be viewed as an indicator of bank performance because it can be uniquely linked to structural indicators of bank behaviour.

6.4.3 Evolution of market power and its determinants

A major advantage of the proposed approach to estimating the Lerner Index is that it gives a better understanding of the evolution of competition over time. In this way, we can then evaluate the factors that determine the individual banks' exercise of market power. These factors could be bank-specific, structural, regulatory or macroeconomic in nature. In the ensuing discussion, we catalogue and motivate the choice of possible determinants of market power in the Zambian commercial banking sector. These variables have theoretical basis as well as empirical support both in developed and developing countries.

Firstly, we control for the structure of the market in which banks operate, depicted by the Hirschman-Herfindahl index (HHI) derived from gross loans.[39] The theoretical rationale for including a measure of market structure is that an individual bank exerts greater influence over the market price relative to marginal cost, thereby increasing the mark-up (Beighley & McCall, 1975; Cowling & Waterson, 1976). This impact is greater the more dominant the bank is and how other banks respond to this dominance. Therefore, we hypothesise that HHI has a positive effect on the Lerner Index.

The second variable we consider is bank credit risk$(RISK)$, defined earlier as the ratio of nonperforming to total loans. Banks' exposure to high credit risk could manifest itself in deterioration of the credit portfolio. To avoid incurring risk, banks may pre-screen their customers and choose to lend o less risky borrowers, even at high interest rates (Stiglitz & Weiss, 1981). Commercial banks may also increase the proportion of risk free assets such as Treasury securities thereby reducing risk-premiums (Tovar, Jaramillo, & Hernandez, 2007). When banks opt to increase their risk exposure, the price-cost margins tend to be higher as banks cover up for the potential loss in revenue arising from default risk by raising their prices relative to marginal cost. However, when the proportion of toxic assets is lower, margins tend to decrease, in which case market power is weaker. Therefore, depending on which of these factors dominates, the effect of the credit risk variable on the Lerner Index cannot be determined a priori.

---

[39] As a robustness check, alternative measures of market structure were introduced. Specifically, the HHI based on total assets and deposits and four-firm concentartion ratio were used in place of the loans based HHI .Estimation with the alternative definitions of HHI produced insignificant and/or wrong signs for coefficients while using the four-firm concentration ratio led to significant loss of observations. Therefore we retained the loans based HHI in the regression.

The next variable considered captures regulatory intensity on banks' exercise of market power. Banks are required by law to hold a certain proportion of risk-weighted assets as regulatory capital. Since banks thrive by engaging in risk lending, minimum capital requirements serve to ensure that banks have a sufficient capital buffer to cover liabilities in an event of bank failure. However, higher regulatory capital requirements could potentially endanger competitiveness by increasing market power. Thus, to capture the impact of regulatory capital requirements (regulatory burden) on market power, we include the capital adequacy ratio $(CAPRATIO)$. It is expected that the intensity of regulatory burden would induce banks to raise their margins in order to build up a sufficient revenue buffer necessary for solvency. Thus, a positive sign is expected on $CAPRATIO$.

Cost inefficiency in banking is often associated with high mark-ups because banks tend to mask their operating inefficiency through wide spreads, the cost of which is borne by customers. In fact, some have interpreted large mark-ups as indicators of cost inefficiency (Vera, Zambrano-Sequin, & Faust, 2007). Thus, when cost inefficiency is a binding constraint, this leads to high market power and may be exacerbated by agency problems. However, it is also possible that banks' high price-cost margins could move in tandem with better cost efficiency performance, mainly because efficient banks are able to contain costs and therefore post wide mark-ups. This is the basis for the efficiency-structure hypothesis which submits that cost efficient banks tend to be more profitable than less efficient ones because of better management performance. For these reasons, the relationship between cost efficiency and mark-ups is unclear.

Another important performance indicator for Zambian banks is the proportion of other income to total assets $(OITASS)$. In the previous chapter, it was established that the degree of competition with respect to fee and other income is defined by oligopolistic conduct. We have also observed that Zambian banks generate 16.8 percent of total income in form of fees, commissions and foreign exchange gains. Therefore, banks with better non-interest revenue performance would exert greater market power, and may use this as an entry barrier. A positive coefficient is expected on the variable $OITASS$. We also control for the ratio of interbank deposits to total customer and short-term funds to account for diversity of bank funding $(DEPMIX)$ on banks' exercise of market power. We conjecture that banks with a high proportion of interbank deposits relative to total deposits will have a low mark-up depicting weaker market power. A negative coefficient is therefore expected on this variable.

Price-cost margins have also been known to vary with the macroeconomic conditions and the monetary policy rule. For instance, during a recession, mark-ups tend to decrease and increase in boom times. Received evidence shows that procyclicality in mark-ups is especially prevalent in manufacturing industries. For the financial services industries, and the banking sector in particular, only limited evidence exists. Small (1998) found that mark-ups in both manufacturing and financial services sector in the United Kingdom were procyclical, indicating that they were high during boom times and when the economy was facing a downturn, mark-ups narrowed. However, Carbo´, et al. (2003) argue that buoyant economic growth and a stable macroeconomic environment tend to negatively affect prices and costs, although the extent to which these variables are influenced may be significantly different. Further evidence by Toolsema (2004) suggests that procyclical monetary policy (increase in policy rate in good times and decrease in bad times) reinforces the countercyclical movement of the Lerner Index, indicating that there is an inverse relationship between monetary conditions and market power. These studies show that the relationship between market power on one hand and the business cycle or monetary policy stance on the other, is ambiguous.

For the present study, the sample period includes some years of economic turbulence as well as episodes of relative macroeconomic stability. Therefore, it is expected that these macroeconomic phenomena would significantly alter banks' pricing and cost behaviour. To evaluate the effects of the business cycle and macroeconomic environment on market power, we include the rate of inflation $(INFLATION)$ as an indicator of cyclicality. The rate of inflation is also an important indicator of broad macroeconomic uncertainty.[40] We expect a positive coefficient on $INFLATION$, suggesting that a high rate of inflation induces banks to increase the price of their bank products whilst cutting down on operating costs to remain competitive. We also include the 91-day Treasury bill rate (TBR) as a measure of monetary policy conditions given that the central bank uses auctions of Treasury securities to raise funds for the government and also as a monetary policy tool for mopping up excess liquidity from the banking system. Therefore, a positive coefficient is expected on TBR.

In view of the foregoing, we estimated Equation (6.9) below to assess the determinants of market power in the Zambian banking sector

[40] In Zambia, the exchange rate is also widely viewed as an important signal of macroeconomic uncertainty. However, its inclusion in the regression yielded poor results, see discussion below. Therefore, we posit that given the strong passthrough effects, see for instance, Mutoti (2006) the rate of inflation adequately captures movements in the exchange rate.

$$\begin{aligned}\ln\left(LI_{it}\right) = {} & \alpha_o + \alpha_1 \ln\left(HHI_{it}\right) + \alpha_2 \ln\left(EFF_{it}\right) + \alpha_3 \ln\left(CAPRATIO_{it}\right) \\ & + \alpha_4 \ln\left(RISK_{it}\right) + \alpha_5 \ln\left(OITASS_{it}\right) + \alpha_6 \ln\left(DEPMIX_{it}\right) \\ & + \alpha_7 \ln\left(INFLATION_t\right) + \alpha_8 \ln\left(TBR_t\right) + v_t \end{aligned} \qquad (6.9).$$

The variables in Equation (6.9) are respectively, Lerner Index (LI), Hirschman-Herfindahl index (HHI), a measure of business cycle and/or macroeconomic policy stance $(INFLATION)$, bank credit risk $(RISK)$, cost efficiency index (EFF), regulatory capital intensity $(CAPRATIO)$, a measure of deposit mix $(DEPMIX)$, a revenue scaling factor $(OITASS)$. Finally, v denotes a random error term and as before, i denotes observation per bank while t is the time of observation in quarters.

6.4.4 Sample and data

As in the preceding chapters, all operating commercial banks between 1998 and 2006 were included in the estimations. Due to entry and mergers, this resulted in unbalanced sample with 359 observations. Since most variables have already been defined in the previous sections, here we only discuss the newly introduced ones (see Table 6.1).

Table 6-1: Variables used in estimating the Lerner Index of market power

Variable	Variable Name	Description and measurement	Mean	Median	Std Dev.	Source
EFF_i	Cost efficiency	Bank-specific cost efficiency score	0.917	0.950	0.099	Stochastic frontier estimates
$DEPMIX_i$	Deposit mix	Share of interbank deposits in total deposits	0.233	0.044	3.341	BS
p_i	Price of bank output	Total individual bank revenue/total bank assets Total revenue	0.022	0.021	0.010	IS and BS
$P = \frac{\sum_{i=1}^{N} p_i}{N}$	Market Price of output	Average of all bank-level output prices	0.022	0.023	0.001	IS and BS
MC_i	Bank-specific marginal cost	Estimated from the translog cost function	0.011	0.010	0.003	Translog cost function estimates
$LI_i = \frac{p_i - MC_i}{p_i}$	Bank-specific Lerner Index	Bank-specific Lerner Index of market power	0.509	0.528	0.021	Estimates of price and marginal cost
$LI = \frac{P - MC}{P}$	Industry- average Lerner Index	Industry average Lerner Index of market power	0.507	0.511	0.141	Estimates of price and marginal cost
$s_i = \frac{LOANS_i}{\sum_{i=1}^{N} LOANS_i}$	Loan market share	Share of individual bank's loans in total loans	7.438	2.871	9.553	BS
$HHI_i = \sum_{i=1}^{N} s_i^2 \times 100$	Hirschman-Hirfindahl index	Market structure concentration index	1970.818	2000.929	182.584	BS

Note: IS – Income statement, BS – Balance Sheet
Source: Bank of Zambia (BoZ) and author's own computations

6.5 Empirical results of market power

Empirical results of the Lerner Index were obtained through joint estimation involving a system of equations given by (6.4) and (6.5) by applying Zellner's (1962) ISURE approach (see Appendix V for parameter estimates of the related cost function). The estimated parameters of the cost function pass the diagnostic tests, with plausible coefficients for the key variables. Most of the coefficients are also significant and of expected signs. These results were used to calculate the bank-specific marginal cost defined by Equation (6.6) which was then used in conjunction with the output price to estimate the corresponding Lerner Index for each bank $\left(LI_i\right)$ as given by Equation (6.7). Averaging across all banks yields the industry level Lerner Index $\left(LI\right)$, depicted by Equation (6.8).

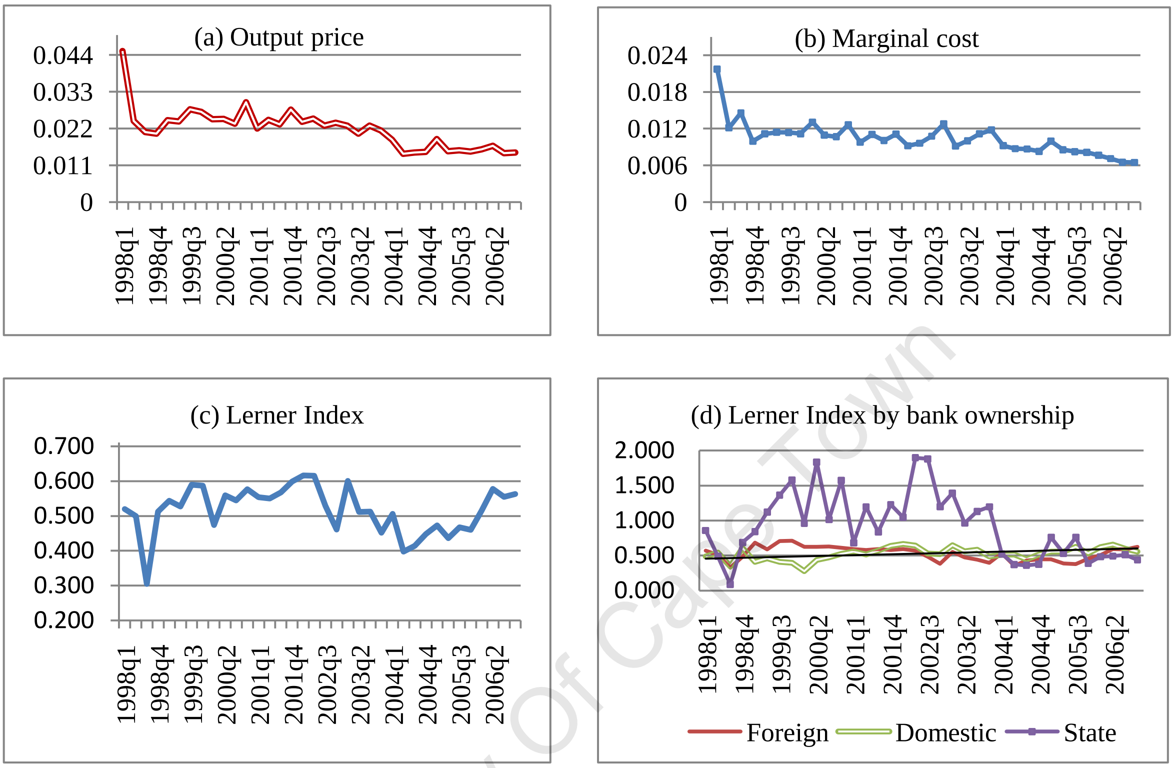
6.5.1 Characteristising the Lerner Index and its evolution

The evolution of the approximate output price, marginal cost and the corresponding Lerner Index are given in Figure 6.1 below. Panel (a) of Figure 6.1 shows that output price had been on the decline throughout the study period, except for a small spike in early 2005. From 2005, bank output price exhibited some degree of stability. This picture is consistent with developments in the banking sector which illustrate a drastic fall in yield rates on securities and a slight decline in bank lending rates. Accordingly, bank interest revenue decreased. A similar observation was made regarding non-interest income, particularly after the implementation of the broad based interbank foreign exchange market (IFEM) system in July 2003, which induced some stability in the exchange rate. In turn, this led to a diminution of foreign exchange gains and as a result non-interest income dipped. Prior to July 2003, banks earned approximately 25.8 percent of total revenue through foreign exchange transactions, shored up by a rapidly depreciating currency. However, this share shrunk to 17.4 percent after the implementation of the IFEM system.

The evolution of banks' marginal cost (Panel b) followed a similar trend as that of bank output price, underpinning the banks' quest to realign their cost structure with slowing growth in revenues from both traditional and non-traditional sources. The reduction in marginal cost accelerated from around 2003 reflecting a decrease in operating and financial costs. This trend continued throughout the remainder of the sample period. The net effect of the decrease in the price of bank output and marginal cost does not necessarily translate into a lower mark-up. Rather, it depends on which one falls faster.

Over the study period, the fall in price of bank output was less rapid compared with the decline in marginal costs. As a result, the Lerner Index trended upwards for most part of the sample period except between 2003q1 and 2004q4 when marginal cost decreased less proportionately than the fall in output price. This can be seen in Panel (c) of Figure 6.1. For the full sample period, the average Lerner Index was estimated to be 50.9 percent, indicating that banks priced above marginal cost by more than 50.0 percent. However, the mark-up decreased from an average of 52.6 percent between 1998 and 2001 to an average of 49.6 percent between 2002 and 2004. Although the Lerner Index bottomed out slightly from about 2005 and 2006, it remained below the pre-2002 level, averaging 49.4 percent over the last two years of the study period.

Figure 6.1: Output price, marginal cost and the Lerner Index (industry average)



Source: Author's own computations from BoZ data

The above analysis shows that between 1998 and 2001, Zambian banks enjoyed greater latitude in setting prices. Therefore, during this period, the Lerner Index was driven more by the high price of bank output than by falling costs. On the other hand, from 2002 onwards the banks' marginal costs decreased precipitously mainly due to a fall in deposit interest rates. At the same time, the price of output also decreased, as banks' lending rates eased in line with a fall in yield rates on Treasury securities. Consequently, the Lerner Index declined marginally, implying a slight decrease in market power over this period.

The estimates of market power based on the Lerner Index corroborate the evidence adduced by the results of the PR approach on the H-statistic, which showed that Zambian banks operated in an imperfectly competitive environment defined by oligopolistic conduct. This oligopolistic behaviour may be due to risk aversion or inadequate predatory strategies that prevented a majority of banks from engaging in intense competition. This is more compelling for state banks (see Panel (d)) which, for reasons of poor credit risk screening mechanisms shied away from engaging in further risky lending. Instead, and supported by soft budgets to cushion them from asset deterioration, they accumulated risk-free securities to boost their revenues. This is evidenced by the relatively higher Lerner Index compared with other bank subgroups. Clearly, the greater exercise of market power by public sector banks was a manifestation of these soft budgets which created an uneven competition platform with other banks.

In view of the above, we could argue that public ownership of commercial banks appeared to have had a strong relevance for Zambia as an entry barrier. As these guarantees were eased, we observe a steep reduction in the Lerner Index from about 2004, even to the point of converging with estimates for other bank categories. Therefore, taking the Lerner Index as an indicator of competitiveness, the evidence given by Figure 6(d) shows that competition among private and foreign banks was tighter while public sector banks operated as a monopoly mainly due to state incentives and guarantees.

6.5.2 Market structure and other determinants of market power

One of the main objectives in this chapter was to study the determinants of market power. To execute this task, the Lerner Index was regressed on measures of market structure, regulatory and macroeconomic variables and bank-specific factors. In

Table 6-2 below we present results of the fixed effects model for the explanatory factors of market power.[41] To control for potential heteroscedasticity, the estimation was conducted using robust standard errors.

[41] The regression equation for the determinants of market power was estimated in the context of a static panel, which does not assume endogeneity of the expanatory variables. If, on the other hand, we suspect that the error term is correlated with any of the explanatory variables, then the system GMM estimation will be the appropriate methodology, in which case the lagged Lerner Index could be used an the instrument.

Table 6-2: Determinants of market power in Zambian banking sector

	Coefficient	Parameter	t-statistic	p-value
Intercept	α_0	-6.166	-2.348	0.019**
$\ln(HHI)$	α_1	0.869	2.332	0.020**
$\ln(EFF)$	α_2	0.585	2.385	0.018**
$\ln(CAPRATIO)$	α_3	0.130	2.011	0.045**
$\ln(RISK)$	α_4	-0.041	-1.824	0.069*
$\ln(OITASS)$	α_5	0.216	4.184	0.000***
$\ln(DEPMIX)$	α_6	-0.055	-2.117	0.035**
$\ln(INFLATION)$	α_7	-0.232	-3.040	0.003***
$\ln(TBR)$	α_8	0.164	3.633	0.000***
σ_u		0.331		
σ_ε		0.366		
ρ		0.449		
No. of Obs.		359		
Wald $\chi^2(8)$		10.730		
p-value		0.000***		
R^2		0.180		
Hausman test		27.390		
p-value		0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent
Source: Author's own calculations based on BoZ data

The Wald-statistic for model adequacy is statistically significant at 1 percent. However, the independent variables explain only 18 percent of the Lerner Index. Given the nature of our sample, this appears rather low. Nonetheless, it is not uncommon for most panel data estimations and cannot be used to authenticate the research findings. Instead, the significance of individual coefficients is more informative in making a case for the usefulness of the results as is evident in Table 6.2 . Details of the these findings are discussed below. The results are indicative of the strong effect of market structure on the Lerner Index power, as depicted by the positive and highly significant (at 5 percent) coefficient on the Herfindahl Hirschman index (HHI). The implication of this result is that market structure is an important factor in explaining banks' market power. This finding is consistent with theoretical predictions and renders support to the assertion that firms operating in concentrated markets tend to exercise market power.

The effect of cost efficiency (EFF) is also robustly significant and of the expected positive sign. This finding deserves special mention in light of the ambiguity in the market power- efficiency relationship discussed earlier. The intuition for this result is that conceptually, cost efficient banks have the ability to exert market power in the Zambian banking sector. In the context of the Zambian banking market, it may be the case that the estimated market power was driven by efficiency gains, as propagated by the ESH theorists. This argument is partly supported by the relatively lower level of cost inefficiency estimated at 8 percent for all banks. This may suggest that most efficient banks could be using cost efficiency as a barrier to competitive behaviour by strengthening their position in the market. Indeed, banks that better manage their productive resources are also able to achieve significant market shares and strategically reposition themselves by posting profits generated from high mark-ups.

The credit risk variable $(RISK)$ has a negative but weakly significant and small coefficient. A unit increase in the ratio of impaired assets to gross loans reduces the size of the mark-up by only 0.4 %. At first glance this finding appears counterintuitive. However, it must be interpreted as follows. Zambian banks have shied away from extending loans to a majority of Zambians, causing many to be shut out of the credit market. Instead, commercial banks have opted for much safer Treasury securities or accumulation of excess reserves (Oxford Policy Management, 2007). In view of this, the risk of default is much less, which means that the role of risk variable in driving market power is weakened. Even when banks have extended credit to the private sector, a majority of these banks (largely foreign owned) have employed robust screening techniques before loan disbursement. As a result, the ratio of NPLs to gross loans is smaller for this category of banks relative to say, public sector banks.[42] Thus, by pre-screening their customers foreign owned banks are able to trade low risk for a small amount of rent. Therefore, the negative parameter estimate on the risk variable must be viewed in the context of the high level of risk averseness which characterises the Zambian banking sector. Furthermore, by establishing lending relationships with their customers, banks acquire

[42] From the available data, of the three categories of banks, namely foreign owned, domestic private and public owned banks, the latter had the highest proportion of non-performing loans (NPLs) relative to gross loans. At an average of 34 percent of gross loans, NPLs for public banks was double that for domestic private banks (18 percent) because this group of banks does is not exemplary with screening loan applicants. Therefore, public owned banks tend to load the risk of default and other charges on the loan rate, thereby intensifyng the degree of market power. On the other hand, foreign banks which boast of better screening techniques had the lowest proportion of NPLs of only 6.8 %. Therefore, it is clear to see that conservative lending, especially by foreign banks, resulted in lower estimate of market power as Panel (d) of Figure 6.1 illustrates.

privileged information their borrowers thereby reducing frictions in the credit market through lock-in effects (Moshe, Kristiansen, & Bent, 2005). This also tends to weaken market power in the banking sector.

As expected, the effect of regulatory capital $(CAPRATIO)$ is positive and statistically significant at 5 percent. This means that well capitalised banks tend to exercise greater market power by virtue of their strength and reputation which manifests itself in capturing a large market share as these banks are deemed safer. Therefore, in the context of the capital buffer theory, this result highlights the fact that banks build up capital to hedge against possible insolvency.[43] Controlling for diversity in revenue sources, the study shows that banks with a greater proportion of other income (fees, commissions, foreign exchange gains, etc.) use this as a device for perpetuating market power. The parameter estimate of $OITASS$ is positive and significant at 1 percent level. Therefore, we submit that including a measure of non-interest revenue in the market power regression helps address the important role other sources of income play in the Zambian banking sector. As demonstrated in preceding discussions, Zambian banks generally continue to charge high fees and commissions and generate significant amount of revenue through foreign exchange transactions. The same can be said about the effect of the TBR on the Lerner Index. Since Zambian banks have historically enjoyed buoyant revenues from investing in Treasury securities, this is captured by the coefficient on the TBR, which was found to be positive and statistically significant at 5 percent.

The negative coefficient on the inflation variable indicates that banks operating under conditions of macroeconomic uncertainty tend to enjoy relatively less market power.[44] Therefore, the main prediction from the analysis is that *ceteris paribus*, bringing down the rate of inflation and improving the economic environment has the possibility of creating incentives for increased market power in the Zambian banking industry. Comparatively, Aboagye, et al. (2008) also found a negative impact of inflation on market power for Ghanaian banks.

[43] It is worth noting that the relationship between capital requirements and bank stability is a subject of ongoing debate, with no conclusive evidence of the direction of causality (Rochet, 1992).
[44] An alternative predictor of macroeconomic conditions would be the gross domestic product (GDP). However, quarterly GDP data for Zambia are unavailable. Therefore, overall macroeconomic stance is proxied by the rate of inflation rate on the premise that macroeconomic uncertainty is inimical to economic growth and therefore high inflation would reasonably approximate deterioration in economic conditions.

The impact of deposit mix $(DEPMIX)$ on the Lerner Index was found to be negative and significant. This result shows that banks with significant recourse to interbank funds suffered a reduction in the relative mark-up. For such banks, short-term interbank borrowing constitutes a fundamental source of funding.

6.6 Summary of findings and comparison with previous studies

In this chapter, we have shown that Zambian banks exercised market power between 1998 and 2006. The estimated average Lerner Index for the full sample was 50.9 percent. The mark-up increased steadily from 1998 until 2001 before petering out in subsequent years.

The findings also indicate that state banks operated as a monopoly, mainly due to soft budgets, which sustained their market power by keeping costs artificially low. The easing of guarantees and other operating incentives levelled the playing field somewhat, thereby narrowing the divergence in the price-cost mark-up between public sector banks and other bank categories. The results also show that bank-specific and environmental factors are important in explaining banks' exercise of market power.

Broadly, these results are comparable with the investigation of market power in previous studies, especially for developing countries. Based on the estimate of the Lerner Index, the overall implication of our findings is that Zambian banks exercised greater market power than other studies have shown. For example, in developing countries, Aboagye, et al. (2008) reported smaller (and in some cases negative) Lerner indices. According to Solis and Maudos (2008), a negative Lerner Index may be evidence of 'super-competition'. Further results by Tovar, et al. (2007) showed that the nominal Lerner Index estimated from the loans market averaged between 20 and 40 percent. In developed countries, Fernandez de Guevara, et al. (2005) and Fernandez de Guevara & Maudos (2007) also found lower estimates of the Lerner Index than those obtained in our study.

Regarding determinants of market power, our results generally confirm those from previous studies. For example, Aboagye, et al. (2008) and Fernandez de Guevara, et al. (2005) showed that cost efficiency is significant in explaining market power, indicating that most efficient banks enjoy higher mark-ups, driven perhaps by lower marginal costs. The authors also find a case for lower inflation in increasing the Lerner Index. Tovar, et al. (2007) argue that the impact of low inflation is pronounced when market power is measured by a real Lerner index. However, our results are at variance with Tovar, et al. (2007) on the role of credit risk. Our findings show that banks with an impaired credit portfolio do not raise their price-cost margins to cover up for the risk. Unlike Moshe, et al. (2005) and Tovar, et al. (2007) our results indicate that the level of concentration is a robust predictor of market power, counteracting our earlier hypothesis that high concentration levels may not be sufficient to explain competition in the Zambian banking sector.

Chapter Seven

Conclusions and policy implications

7.1 Introduction

This research has taken a systematic and an in-depth analysis of the performance of commercial banks in Zambia in the post-liberalisation era. In particular, due to data constraints, we focussed on the period after the banking crisis, that is, the sample covering 1998-2006. The study was presented in seven chapters, including the introductory one. Of the seven chapters, three were devoted to empirical analysis of inter-related research issues. Chapter two, which followed the introductory section looked at the financial landscape in Zambia and presented some performance indicators of the banking sector. Chapter three was an overview of some of the microeconomic theories of banking, with a particular focus on those related to the issues addressed in the thesis. Chapters four, five and six provided an empirical investigation of the main themes of the study as presented below.

The first theme, examined in chapter four, relates to banks' efficiency performance and the factors that explain cost inefficiency in the Zambian banking industry. This issue is central to continued viability of the banking industry and stems from a broader policy concern that Zambian banks operate inefficiently. The second issue, addressed in chapter five, is an evaluation of competitive conditions in the Zambian banking system. Based on reduced form revenue equations, we estimated the degree of competition relating to the market structure which characterised the Zambian commercial banking sector. Chapter six addressed an important element of banks' conduct in terms of pricing and cost decisions, namely, exercise of market power approximated by the Lerner Index. The Lerner Index is an important concept in banking markets which exhibit oligopolistic characteristics. Since the Lerner Index can be estimated for individual banks and across time, we used this to assess the evolution and intensity of competition in the banking industry. This provided a robustness check to the results of chapter five.

7.2 Main findings

The main objective of the study was to provide an assessment of the three pillars of bank performance and conduct in the Zambian commercial banking system in the context of a liberalised environment. Specific research objectives were articulated in each of the empirical chapters to serve as a guide for testing of the stated hypotheses. The main findings from each of the issues tackled are discussed below.

7.2.1 Cost efficiency and its determinants

Cost efficiency is an important aspect of performance monitoring of the banking industry and it is one area bank regulators focus on when conducting an analysis of the viability of the industry. This study went beyond the financial ratio analysis often employed by managers and regulators to benchmark performance. In this study, one of the objectives was to estimate cost efficiency (inefficiency) of commercial banks and its evolution over time during 1998-2006. A common stochastic frontier approach was estimated jointly with the inefficiency model in a single step approach in order to simultaneously investigate the level of cost inefficiency and the factors that affect it. In this regard, we evaluated the importance of both bank-specific and environmental variables as determinants of efficiency performance.

The main findings from the analysis are:

i. During the period of investigation, an average Zambian bank deviated from the best-practice frontier. The average cost efficiency level for all banks was 92 percent. This estimate suggests that an average bank could have incurred about 8 percent less in actual costs if it had been utilising its input resources in an optimal fashion.

ii. The results reveal that the most cost efficient bank had the efficiency level of 96.2 percent, implying that to match the best-practice frontier estimate, this bank could have cut its costs by 3.8 percent. At the polar end of the efficiency ranking was the least efficient bank with cost inefficiency level of 30.8 percent.

iii. The investigation of the relationship between bank ownership structure and cost efficiency indicated that government owned banks were better positioned relative to other bank catergories, that is, domestic private and foreign owned banks. However, this performance must be viewed in the context of guarantees and other incentives extended to public sector banks aimed at ridding the balance sheet of toxic assets for these institutions. In all likelihood, these soft budget provisions propped up cost performance of the public owned institutions, thereby raising the level of efficiency beyond what would otherwise have been the case in the absence of these support mechanisms.

iv. The study revealed a sustained uptake in the level of efficiency of domestic private banks, surpassing even that of foreign owned banks in the latter period of the sample. This improvement came against a backdrop of dynamic positive changes in the broader macroeconomic enviornment. This means that the domestic private banks responded favourably to conditions induced by favourable macroeconomic developemnts. The leap in cost efficiency was also underpinned by the merger in 2004 between two of the small domestic private banks. This merger stimulated the average efficiency level well above that which existed before the merger was authorised.

v. Results from the regression of the determinants of cost inefficiency show that regulatory policy did not exert a distortionary impact on banks' efficiency peformance. However, output quality measured by bank credit risk exacerbated cost inefficiency of Zambian banks. A high proportion of assets held in Treasury scurities also had a detrimental effect on efficiency performance.

vi. Generally, the study has established that instability in the macroeconomic environment created uncertainty in the banking market. As a result, banks failed to perform efficiently. This was evidenced by the positive impact of the rate of inflation on cost inefficiency.

7.2.2 Structure and competition in the banking system

This study estimated a competitiveness index in order to address the ambiguity between concentration and competition in the banking industry. This test also sought to dispel claims that the dominance of foreign owned banks and continued state ownership of some banks may have stifled competition in the banking system. The broad conclusions from this analysis are summarised as follows:

i. The study finds evidence of oligopolistic behaviour in the Zambian banking sector. The estimated $H-$statistic which denotes competitiveness was positive and significantly different from zero and unit across all revenue specifications. This finding indicated that Zambian banks earned their revenue in a monopolistically competitive environment. Consequently, the results rejected any form of perfect competition and monopoly as possible explanations of market behaviour. The results were also robust to different definitions of bank classes.

ii. The magnitude of the $H-$statistic varied among different bank subgroups and revenue measures. Based on the estimates from total revenue, the intensity of compeition was greater for foreign owned banks than among domestic banks. The converse was the case using interest revenue. Domestic banks had the largest $H-$statistic compared with foreign owned banks. Results of the estimated competitive index from both total and interest revenue equations suggested that small banks were inclined to behave more competitively than large counterparts. This may be attributed to the small banks' quest to capture the elusive market share in the banking system from the large competitors.

iii. The validity of the test for long-run equilibrium gave further impetus to the results, suggesting that the reduced form revenue equations correctly captured the state of competition in the Zambian banking system.

7.2.3 Banks' exercise of market power: evolution and sources

The chapter on the banks' exercise of market power and its determinants extended the analysis of the preceding discussion by evaluating the evolution of pricing and cost decisions at bank level and over time. The specific focus of the study was the measurement of market power proxied by the Lerner Index using information on output price and marginal cost derived from a translog cost function. Knowledge of banks' exercise of market power is of particular relevance for regulatory policy because it provides information on the actual behaviour regarding the banks' influence over price and output in the banking sector. Banks also need to have a clear view of their market position to ensure that they are not the focus of monopoly regulators. The main conclusions drawn from this chapter are presented below:

i. Analysis of results shows that the banks' output price was more sticky than the decrease in costs, especially during 1998-2001. This translated into an average Lerner Index of about 50.9 percent implying a relative mark-up of price over marginal cost in excess of 50 percent. This finding provides evidence that for the industry as a whole, market power was no where near the monopoly case and was still further away from the polar end of perfect competition.

ii. The Government's support mechanisms to state owned banks perpertuated the exercie of market power by these institutions by keeping costs artificially lower relative to the price of aggregate output. Thus, it was observed that the estimated Lerner Index of public sector banks was larger than for domestic private and foreign owned banks. However, this conduct dissipated in the later years as government tightened its grip on inflation, causing a cessation in the provision of balance sheet incentives to public owned banks. On the other hand, domestic private and foreign owned banks exhibited similar degree of market power. For both cases, the estimate of the Lerner Index was around 53 percent, marginally higher than the sample average.

iii. The results also show that the level of concentration had a positive impact on the Lerner Index. This indicated that market dominance may have been reflected in greater influence over price of output. The response of the Lerner Index to the increasing level of concentration was found to be around 0.9 percent. Moreover, cost efficient banks posted solid mark-ups. This means that by managing their productive resources in an efficient manner, banks gained control over price and costs and therefore acquired greater market power. The study also established that banks with substantial share of non-intermediated sources of revenue enjoyed greater market power.

iv. The study also revealed an important issue relating to the response of banks' markups to macroeconomic conditions. It was observed that market power in the Zambian banking system is affected by the degree of macroeconomic stance, high in bad times and lower under stable macroeconomic conditions. This finding fits nicely with the estimated parameters on the rate of inflation in the revenue equations for the Panzar-Rosse competitiveness measure which showed that instability in the macroeconomic environment constrains the banks' competitive conduct by dampening the sensitivity of revenue to macroeconomic currents.

v. Regulatory intensity fosters market power. This result, consistent with the capital buffer proposition suggests that highly capitalised banks use their capital strength as a premium on price-cost margins. In this streak, raising capital standards beyond the existing threshold may endanger competitiveness in the banking sector. In contrast, banks with a high proportion of impaired assets exerted a weak dampeding influence on output price. The result follows from the credit rationing model which posits that through better screening of borrowers banks are only able to make safe loans, thereby reducing risk. As a result, this entrenches their position in the market.

7.3 Policy implications and recommendations

Results derived in this thesis provide some valuable policy lessons. These are broadly summarised as follows:

i. An important factor that policy makers should take note is that, while the deregulation of the banking sector was introduced to remove constraints in the market and increase bank efficiency, the distortions caused by years of administered interest rates and excessive government intervention in the banking industry continue to persist. Therefore, banks could record additional efficiency gains through further deepening of reforms and dealing with some of the outstanding performmance impediments, such as instability in the macroeconomic environment. Specifically, containing inflation will create incentives for improved bank performance.

ii. Although the results from the analysis of competition and market power provide evidence of departure from monopolistic behaviour, there is room for exploiting possibilities to strengthen the degree of competition and dampen market power. Therefore, regulatory authorities should design measures aimed at creating further incentives for enhanced competitiveness in the banking sector. In particular, the regulatory authorities should endeavour to create an enabling environment for contestability in the banking industry, for example contuining with the open policy of allowing foreign as well as domestic bank entry into the sector. This will intensify competition and propagate efficiency gains across the banking market. Furthermore, the policy of setting up a credit reference bureau (CRB) should be welcomed in that it will help lower informational costs related to screening and monitoring of borrowers which in turn will enhance the quality of loans by reducing the probability of default.

iii. Where banks have been able to sustain their risk aversion by holding a large poroportion of risk-free Treasury securities, the Government should step up its efforts of reducing its borrowing requirements from the banking sector. This will inevitably remove the cushionining effect banks currently enjoy from holding large amounts of Government debt and stimulate them to innovate and develop better mechanisms of coping with risk related to private sector lending.

iv. At the same time, government should refrain from providing gaurantees and other incentives that distort the operating environment for commercial banks. Previously, implicit and explicit state support extended to the Zambia National Commercial Bank (ZNCB) kept the bank's costs artificially low, allowing it to operate as a monopoly to the detriment of other banks in the industry. Therefore, it is gratifying to note that government decided to relinquish its direct involvement in the banking sector by concluding the privatisation of ZNCB. This is critical to achieving the highest degree of impartiality in the credit decision-making process and creating a competitive level playing field in the industry.

v. Past legislative and regulatory changes have certainly benefited the efficiency of Zambian banks and if sustained could improve contestability as well. Therefore, looking ahead, there are economic reasons to suggest the importance of continuing to implement policies that promote efficiency and competition in the banking sector and the financial services industry in general. In particular, the authorities should place emphasis on the legal and institutional framework for resolving problem credit situations, including setting up a commercial court to expedite recovery of collateral. This is because a relatively high incidence of problem loans remains a daunting challenge for many commercial banks.

7.4 Contributions and limitations of the study

Hitherto, there has been no study to our knowledge that addresses the core elements of banks' performance in an integrated framework. Thus, this thesis presents an initial effort at quantifying the performance of Zambian banks since the end of the crisis that affected the industry in the mid-1990s. The structure of the empirical sections enabled us to isolate the key elements of the banks' behaviour, inter alia cost efficiency, market structure and banks' mark-ups, all important for regulatory policy.

By conducting an in-depth investigation of these pertinent research issues, the thesis has extended the frontier of knowledge by providing another dimension for analysing the cost performance and competitive behaviour of Zambian banks. Although the analysis is specific to the Zambian banking sector, the approach taken and results derived can be applied to other developing countries, especially in SSA where banking sector reforms have been implemented. This is the main contribution of the thesis.

However, the study suffers from some limitations as well. The key ones are highlighted below:

i. Failure to make a comparison of banks' performance during pre-reform and post-reform period is a major shortcoming of this thesis. From the results obtained, we cannot say with certainty whether or not there has been a change in the level of competition and market power and indeed cost efficiency stemming from the implementation of financial and regulatory reforms.

ii. In the course of conducting this research, it has become clear that a major constraint on a more precise study of the issues considered is a shortage of relevant detailed data on product/service specific prices and costs. Certainly, this precluded us from conducting an investigation of the degree of market power in each of the banks' business lines in order to gain a better perspective on banks' price setting and cost behaviour.

iii. Finally, we have not been able to determine the welfare implications associated with the banks' exercise of market power. Consequently, it is premature to state categorically if the estimated degree of market power led to loss of social welfare during the study period.

7.5 Possible areas for future research

In view of the limitations of this study, future research efforts may explore ways of adequately incorporating the implications of risk taking behaviour on the banks' performance. This is particularly relevant given the recent experience of global financial crisis, which has exposed banks' risk taking on the viability of financial institutions and other sectors of the economy.

Related to this, research has suggested that competition in banking leads to more risk taking with potential for bank runs. However, the trade-off between bank competition, risk and financial stability has remained an unresolved empirical issue.

Thus, to the extent that high competitiveness of Zambian banks may lead to instability in the financial sector, and drawing from past experiences bank failures, new research should assess the extent to which increased competition poses a threat to the soundness and stability of the banking system. Furthermore, future research should interrogate whether or not high competition in the banking sector fosters increased access to finance, particularly by small firms.

Bibliography

Abel, A. B., Mankiw, N. G., Summers, L. H., & Zeckhauser, R. J. (1989). Assessing dynamic efficiency: Theory and evidence. *The Review of Economic Studies , 56*, 1-19.

Aboagye, A. Q., Akoena, S., Antwi-Asare, T., & Gockel, A. (2008). Explaining the market power of Ghanaian banks. *South African Journal of Economics , 76*, 569-585.

Adongo, J., Stork, C., & Hasheela, E. (2005a). *Fcators influencing the alternative profit X-efficiency of Namibia's banking sector*. Retrieved June 2007, from http://www.nepru.org.na/index.php?id=247&no_cache=1&file=880&uid=424.

Adongo, J., Stork, C., & Hasheela, E. (2005b). *Measuring alternative profit X-efficiency Namibia's banking sector*. Retrieved June 2007, from http://www.nepru.org.na/index.php?id=126&no_cache=1&file=461&uid=183.

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

Akerloff, G. (1970). Market for 'lemons': Quality uncertainty and market mechanism. *The Quarterly Journal of Economics , 84*, 488-500.

Alencar, L. S., & Nakane, M. I. (2004). Bank competition, agency costs and the performance of monetary policy. *Working Paper Series No.81* . Banco Central do Brasil.

Allen, F., & Gale, D. (2000). *Comparing Financial Systems*. Cambridge: Massachusettes Institute of Technology (MIT) Press.

Allen, F., & Gale, D. (2004). Competition and financial stability. *Journal of Money, Credit and Banking , 36*, 453-480.

Allen, F., & Santomero, A. M. (2001). What do financial intermediaries do? *Journal of Banking and Finance , 25*, 271-294.

Al-Obaidan, A. M. (2008). Market structure concentration and performance in the commercial banking industry of emerging markets. *European Journal of Economics, Finance and Administrative Sciences , 12*, 104-114.

Altunbas, Y., Liu, M.-H., Molyneux, P., & Seth, R. (2000). Efficiency and risk in Japanese banking. *Journal of Banking and Finance , 24*, 1605-1628.

Aly, H. Y., Grabowski, R., Pasurka, C., & Rangan, N. (1990). Technical, scale, and allocative efficiencies in U.S. banking: An empirical investigation. *Review of Economics and Statistics , 72*, 211-218.

Amato, L. (1995). The choice of structure measure in industrial economics. *Quarterly Journal of Business and Economics , 32*, 39-52.

Angelini, P., & Cetorelli, N. (2003). The effects of regulatory reform on competition in the banking industry. *Journal of Money, Credit and Banking , 35*, 663-684.

Ansari, M. (2007). An empirical investigation of cost efficiency in the banking sector of Pakistan. *State Bank of Pakistan Research Bulletin , 3*, 209-231.

Appelbaum, E. (1982). The estimation of the degree of oligopoly power. *Journal of Econometrics , 19*, 287-299.

Aryeetey, E., Hemamala, H., Machiko, N., & Steel, W. (1994). *Financial market segmentation and reforms in Sub-Saharan Africa.* Washington,DC: World Bank.

Ataullah, A., & Le, H. (2006). Economic reforms and bank efficiency in developing countries: the case of the Indian banking industry. *Applied Financial Economics , 16*, 653-663.

Ataullah, A., Cockerel, T., & Le, H. (2004). Financial liberalisation and bank efficiency: a comparative analysis of India and Pakistan. *Applied Economics , 36*, 1915-1924.

Ausunbel, L. (1991). The failure of competition in the credit card market. *American Economic Review , 81*, 50-81.

Bain, J. (1951). Relation of profit rate to industry concentration: American manufacturing 1936-1940. *The Quarterly Journal of Economics , 65*, 293-324.

Baltensperger, E. (1980). Alternative approaches to the theory of the banking firm. *Journal of Monetary Economics , 6*, 1-37.

Banker, R., Charnes, A., & Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science , 30*, 1078-1092.

Battese, G. E., & Corra, G. (1977). Estimation of a production frontier model with application to the pastoral zone of eastern Australia. *Australian Journal of Agricultural Economics , 21*, 167–179.

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.

Bauer, P., Berger, A., Ferrier, G., & Humphrey, D. (1998). Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. *Journal of Business and Finance , 50*, 85-114.

Baumol, W. (1982). Contestable markets: an uprising in the theory of industry structure. *American Economic Review , 72*, 1-15.

Beccalli, E., Casu, B., & Girardone, C. (2006). Efficiency and stock performance in European banking. *Journal of Business Finance & Accounting , 33* (1-2), 245-262.

Beck, T. (2008, June). Bank competition and financial stability: friends or foes? *Policy Research Working Paper,No.4656* . The World Bank.

Beck, T., & Hesse, H. (2009). Why are interest spreads so high in Uganda? *Journal of Development Economics , 88*, 198-204.

Beck, T., Demirgüç-Kunt, A., & Maksimovic, V. (2004). Bank Competition and access to finance: international evidence. *Journal of Money, Credit and Banking , 36*, 627-648.

Beighley, P. H., & McCall, A. (1975). Market power and structure and bank installment lending. *Journal of Money,Credit and Banking , 7*, 449-467.

Belaisch, A. (2003, May). Do Brazilians banks compete? *IMF Working Paper WP/03/113* . International Monetary Fund.

Berg, A. S., & Kim, M. (1998). Banks as multioutput oligopolies: an empirical evaluation of the retail and corporate banking markets. *Journal of Money,Credit and Banking , 30*, 135-153.

Berger, A. N. (2007). International comparisons of banking efficiency. *Financial Markets, Institutions and Instruments , 16*, 119-144.

Berger, A. N. (1995). The profit-structure relationship in banking: tests of market power and efficient structure hypothesis. *Journal of Money, Credit and Banking , 27*, 404-431.

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

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

Berger, A. N., & Humphrey, D. (1991). The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics , 28*, 117-148.

Berger, A. N., Hunter, W., & Timme, S. (1993). The efficiency of financial institutions: a review and preview of research past, present and future. *Journal of Banking and Finance , 17*, 221-249.

Berger, A. N., Leusner, J. H., & Mingo, J. J. (1997). The efficiency of bank branches. *Journal of Monetary Economics , 40*, 141-162.

Berger, A., Leusner, J., & Mingo, J. (1994, August). The efficiency of bank branches. *Working paper no. 94–27* . The Wharton Financial Institutions Centre, Federal Reserve Board.

Berstein, S., & Fuentes, R. (2003). Is there lending rate stickiness in the Chilean banking industry. *Working paper No.218* . Central Bank of Chile.

Besanko, D., & Thakor, A. (1992). Banking deregulation: allocational consequences of relaxing entry barriers. *Journal of Banking and Finance , 16*, 909-932.

Bhattachayra, S., & Thakor, A. (1993). Contemporary banking theory. *Journal of Financial Intermediation , 3* (1), 2-50.

Bhaumik, S. K., & Dimova, R. (2004). How important is ownership in a market with level playing field? The Indian banking sector revisited. *Journal of Comparative Economics , 32*, 165-180.

Bikker, J. A., & Bos, J. W. (2005). Competition and efficiency in banking: international comparisons. *Economic & Financial Modelling* (Autum), 103-148.

Bikker, J., & Haaf, K. (2002). Competition, concentration and their relationship:an empirical analysis of the banking industry. *Journal of Banking and Finance , 26*, 2191-2214.

Bikker, J., & Haaf, K. (2003). Measures of competition and concentration in the banking industry: a review of the literature. *Economic and Financial Modelling* (Summer).

Boitumelo, M., & Valadkhani, A. (2008). *Technical efficiency in Botswana's financial institutions:a DEA approach*. Retrieved from http://www.uow.edu.au/commerce/econ/wpapers.html

Bolt, W., & Tieman, A. F. (2004). Banking Competition, Risk, and Regulation. *IMF Working Paper No. WP/04/11* . International Monetary Fund.

Bottasso, A., & Sembenelli, A. (2004). Does ownership affect firms' efficiency: Panel data evidence on Italy. *Empirical Economics , 29*, 769-786.

Bresnahan, T. F. (1989). Empirical studies of industries with market power. In R. Schmalensee, & R. Willig (Eds.), *Handbook of Industrial Organization* (Vol. 2, pp. 1011-1057). Amsterdam: North-Holland.

Bresnahan, T. F. (1982). The oligopoly solution concept is identified. *Economic Letters , 10*, 87-92.

Brownbridge, M. (1998). Financial policies and banking in Zambia. In M. Brownbridge, & C. Harvey, *Banking in Africa.* New Jersey: Africa World Press, Inc.

Brownbridge, M., & Gayi, S. (1999). Progress, constraints and limitations of financial sector reforms in least developed countries. *Discussion Paper No.7* . Geneva: United Nations Conference on Development and Trade (UNCTAD).

Bruno, G. (2004). A comparative analysis of alternative econometric packages for the unbalanced two-way error component model. *Information Technology for Economics and Management , 4*, 1-18.

Buchs, T. D., & Mathisen, J. (2005). Competition and efficiency in banking: behavioural evidence from Ghana. *Working Paper, WP/05/17* . Washington DC: International Monetary Fund.

Carbó Valverde, S., Humphrey, B. D., & López del Paso, R. (2007). Opening the black box: finding the source of cost inefficiency. *Journal of Productivity Analysis , 27*, 209-220.

Carbó Valverde, S., Humphrey, D., & Rodriguez, F. (2003). Bank deregulation is better than mergers. *International Financial Markets, Institutions and Money , 13*, 429-449.

Carbó Valverde, S., Humphrey, D., & Rodriguez, F. (2003). Deregulation, bank competition and regional growth. *Regional Studies , 37*, 227-237.

Carvallo, O., & Kasman, A. (2005). Cost efficiency of banks in the Latin American and the Caribbean banking systems. *International Financial Markets, Institutions and Money , 15*, 55-72.

Casu, B., & Girardone, C. (2006). Bank competition, concentration and efficiency in the single European market. *Manchester School , 74*, 441-468.

Cebenoyan, A. S. (1988). Multiproduct cost functions and scale economies in banking. *The Financial Review , 23*, 499-512.

Cetorelli, N. (1999). Competitive analysis in banking: Appraisal of methodologies. *Economics Perspectives , First Quarter*, 2-15.

Cetorelli, N., & Gamberra, M. (2001). Banking market structure, financial dependence and growth: International evidence from industry data. *Journal of Finance , 56* (2), 617-648.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research , 2*, 429-444.

Chen, C. (2009). Bank efficiency in Sub-Saharan African middle-income countries. *IMF Working Paper, WP/09/14* . International Monetary Fund (IMF).

Chen, X., Skully, M., & Brown, K. (2005). Banking efficiency in China: Application of DEA to pre- and post-deregulation eras: 1993–2000. *China Economic Review , 16*, 229-245.

Chirwa, E. (2003). Determinants of commercial banks' profitability in Malawi: a cointegration approach. *Applied Financial Economics , 13*, 565-571.

Chirwa, E. (2001). Market structure, liberalisation and performance in the Malawian banking industry. *Research Paper 108* . Nairobi: African Economic Research Consortium (AERC).

Chirwa, E. W., & Mlachila, M. (2004). Financial reforms and interest rate spreads in the commercial banking system in Malawi. *IMF Staff Papers , 51*, 96-122.

Christensen, L. R., Jorgenson, D., & Lau, L. (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics , 55*, 28-45.

Christopoulos, D., & Tsionas, E. G. (2001). Banking economic efficiency in the deregulation period: results from heteroscadastic stochastic frontier models. *The Manchester School , 69*, 656-676.

Čihák, M., & Podpiera, R. (2005). Bank behavior in developing countries: evidence from East Africa. *Working Paper No.WP/05/129* . International Monetary Fund.

Claessens, S., & Laeven, L. (. (2005). Financial dependence, banking sector competition,and economic growth. *Journal of the European Economic Association , 3* (1), 179-207.

Claessens, S., & Laeven, L. (2004). What drives bank competition? Some internantional evidence. *Journal of Money, Credit and Banking , 36* (3), 563-583.

Claessens, S., Demirgüc-Kunt, A., & Huizinga, H. (2001). How does foreign bank entry affect domestic banking markets? *Journal of Banking and Finance , 25* (5), 891–911.

Clarke, G. R., Cull, R., & Fuchs, M. (2009). Bank privatization in sub-Saharan Africa: The case of Uganda commercial bank. *World Development , 37* (9), 1506-1521.

Coccorese, P. (2009). Market power in local banking monopolies. *Journal of Banking & Finance , 33*, 1196-1210.

Corviosier, S., & Gropp, R. (2002). Bank concentration and retail interest rates. *Journal of Banking and Finance , 26*, 2155-2189.

Cowling, K., & Waterson, M. (1976). Price-cost margins and market structure. *Economica , 43*, 267-274.

Crampton, P. (2003, April 7-8). *Competition and efficiency as organising principles for all economic and regulatory policymaking*. Retrieved May 2007, from http://www.oecd.org: http://www.oecd.org/dataoecd/43/26/2490195.pdf

de Luna Martinez, J. (2006). Access to financial services in Zambia. *World Bank Policy Working Paper, WPS4061* . Washington DC: World Bank.

Delis, D. M., Staikouras, K. C., & Varlagas, T. (2008). On the measurement of market power in the banking industry. *Journal of Business Finance and Accounting , 35*, 1-25.

Demsetz, H. (1973). Industry structure, market rivalry and public policy. *Journal of Law and Economics , 16*, 1-9.

Denizer, C. A., Dinc, M., & Tarimcilar, M. (2007). Financial liberalisation and banking efficiency: evidence from Turkey. *Journal of Productivity Analysis , 27*, 177-195.

Dermine, J. (1986). Deposits rates, credit rates and bank capital: the Klein-Monti model revisited. *Journal of Banking and Finance , 10*, 99-114.

Dermine, J. (1984). *Pricing Policies of Financial Intermediaries.* London: Springer.

Dewatripont, M., & Tirole, J. (1994). *The Prudential Regulation of Banks.* Cambridge: Massachusettes Institute of Technology (MIT) Press,.

di Patti, E. B., & Hardy, D. C. (2005). Financial sector liberalisation, bank privatisation and efficiency: evidence from Pakistan. *Journal of Banking and Finance , 29*, 2381-2406.

Diamond, D. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies , 51*, 383-414.

Diamond, D., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy , 91*, 401-419.

Diebold, F., & Sharpe, S. (1990). Post-deregulation bank-deposit-rate pricing: The multivariate dynamics. *Journal of Business and Economic Statistics , 8*, 281-291.

Drakos, K., & Konstantinou, P. (2005). Competition and contestability in transition banking: an empirical analysis. *South-Eastern Europe Journal of Economics , 2*, 183-209.

Dvořák, P. (2005). Rethinking the Monti-Klein model of banking industry: new insights about the separability of loans and deposits decisions. *Discussion Paper No. 2005-138* . Centre for Economic Research and Graduate Education, CERGE-EI.

Eisenbeis, R., Ferrier, G. D., & Kwan, S. H. (1999). The informativeness of stochastic frontier and programming frontier efficiency scores: cost efficiency and other measures of Bank Holding Company performance. *Working Paper, No.99-33* . Federal Reserve Bank of Atlanta.

Elliehausen, G. (1998). The cost of bank regulation: a review of the evidence. *Federal Reserve Bulletin , April*, 1-35.

Estrada, D., & Osorio, P. (2004). Effects of financial capital on Colombian banking effciency. Banco de la Republica de Colombia.

Evanoff, D. D., & Fortier, D. L. (1988). Reevaluation of the structure-conduct-performance paradigm in banking. *Journal of Financial Services Research , 1*, 277-294.

Fahrer, J., & Rohling, T. (1981). Some tests of competition in the Australian housing loan market. *Australian Economic Papers , 33*, 107-119.

Fare, R., Grosskopf, S., & Weber, W. L. (2004). The effect of risk-based capital requirements on profit efficiency in banking. *Applied Economics , 36*, 1731-1743.

Farell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society , 120*, 253-290.

Favero, C., & Papi, L. (1995). Technical efficiency and scale efficiency in the Italian banking sector:a non-parametric approach. *Applied Economics , 27*, 385-395.

Fernandez de Guevara, J., & Maudos, J. (2007). Explanatory factors of market power in the banking system. *The Manchester School , 75*, 275-296.

Fernandez de Guevara, J., Maudos, J., & Perez, F. (2005). Market power in European banking sectors. *Journal of Financial Services Research , 27*, 109-137.

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

Figueira, C., Nellis, J., & David, P. (2006). Does ownership affect the efficiency of African banks? *Journal of Developing Areas , 40*, 37-62.

FinMark Trust. (2006). *Measuring financial access in Zambia, FinScope Zambia 2005: Summary of topline findings.* Lusaka: Government of the Republic of Zambia.

Fischer, K.-H., & Hempell, H. S. (2006). *Regional markets, oligopoly and market power in banking.* Retrieved August 2008, from teaching.coll.mpg.de/econwork/FischerHempell.pdf

Frame, S., & Kamerschen, D. (1997). The profit-structure relationship in legally protected banking markets using efficiency measures. *Review of Industrial Organisation , 12*, 9-22.

Franks, J. R., Schaefer, S. M., & Staunton, M. D. (1997). The direct and compliance costs of financial regulation. *Journal of Banking and Finance , 21*, 1547-1572.

Freixas, X., & Rochet, J. (1997). *Microeconomics of banking.* Cambridge: Massachusettes Institute of Technology (MIT) Press.

Freixas, X., & Santomero, A. M. (2002). An overall perspective on banking regulation. *Federal Reserve Bank of Philadelphia Working paper 02-1* . Federal Reserve Bank of Philadelphia.

Fries, S., & Taci, A. (2005). Cost efficiency of banks in transition: evidence from 289 banks in 15 post-communist countries. *Journal of Banking and Finance , 29*, 55-81.

Fu, X., & Heffernan, S. (2007). Cost X-efficiency in China's banking sector. *China Economic Review , 18*, 35-53.

Fujii, A. (2001). Determinants and probability distribution of inefficiency in the stochastic cost frontier of Japanese hospitals. *Applied Economics Letters , 8*, 807-812.

Gelos, G., & Roldos, J. (2004). Consolidation and market structure in emerging market banking systems. *Emerging Markets Review , 5*, 39-59.

Gilbert, A. R. (1984). Bank market structure and competition: A survey. *Journal of Money, Credit and Banking , 16*, 617-645.

Giokas, D. (2008). Assessing the efficiency in operations of a large Greek bank branch network adopting different economic behaviors. *Economic Modelling , 25*, 559-574.

Government of the Republic of Zambia. (2007). *Statement by the Minister of Commerce Trade and Industry on the partial privatisation of ZANACO*. Retrieved March 2, 2009, from http://www.mibs.gov.zm/index.php?option=com_content&task=view&id=145&Itemid=142

Green, W. H. (2003). *Econometric Analysis.* New Jersey: Prentice Hall.

Grigorian, D., & Manole, V. (2006). Determinants of commercial bank performance in transition: an application of data envelopment analysis. *Comparative Economic Studies , 48*, 497-522.

Gropp, R., Sørensen, C. K., & Lichtenberger, J.-D. (2007). The dynamics of bank spreads and financial structure. *Working Paper Series, No.714* . European Central Bank.

Gruben, W., & McComb, R. (2003). Privatisation, competition and supercompetition in the Mexican commercial banking system. *Journal of Banking and Finance , 27*, 229-249.

Gunji, H., Miura, K., & Yuan, Y. (2009). Bank competition and monetary policy. *Japan and the World Economy , 21*, 105-115.

Gurley, J., & Shaw, E. S. (1955). Financial aspects of economic development. *American Economic Review , 45*, 515-538.

Guzman, M. (2000). The economic impact of bank structure: a review of recent literature. *Economic and Financial Review* (Second Quarter), 11-22.

Hadri, K. (1999). Estimation of a doubly heteroscedastic stochastic frontier cost function. *Journal of Business and Economic Statistics , 17*, 359-363.

Hanak, E. E. (1992). A service-based theory of retail banking. *Journal of Managerial and Decision Economics , 13*, 183-200.

Hanan, T., & Berger, A. (1991). The rigidity of prices: evidence from the banking industry. *American Economic Review , 81*, 938-945.

Hanan, T., & Liang, N. (1993). Inferring market power from times series data: the case of of the banking firm. *International Journal of Industrial Organisation , 11*, 205-218.

Hannan, T. (1991b). Bank commercial loan markets and the role of market structure: evidence from the survey of commercial lending. *Journal of Banking and Finance , 15*, 133-149.

Hannan, T. (1991a). Foundations of the Structure-Conduct-Performance paradigm in banking. *Journal of Money, Credit and Banking , 23*, 68-84.

Hannan, T. (1997). Market share inequality, the number of competitors, and the HHI: An examination of bank pricing. *Review of Industrial Organization , 12*, 23-35.

Hao, J., Hunter, W., & Yang, W. (2001). Deregulation and efficiency: the case of private Korean banks. *Journal of Economics and Business , 53*, 237-254.

Harvey, C. (1973). The control of credit in Zambia. *The Journal of Modern African Studies , 11*, 383-92.

Hauner, D. (2008). Credit to government and banking sector performance. *Journal of Banking and Finance , 32*, 1499-1507.

Hauner, D., & Peiris, S. J. (2008). Banking efficiency and competition in low income countries: the case of Uganda. *Applied Economics , 40*, 2703-2720.

Hausman, J. A., & Sidak, G. J. (2007). Evaluating market power using competitive bechmark prices instead of the Herfindahl-Hirschman index. *Antitrusy Law Journal , 74*, 387-407.

Herrero, I., & Pascoe, S. (2002). Estimation of technical efficiency: a review of some of the stochastic frontier and DEA software. *Computers in Higher Education Economics Review , 15* (1).

Hodgman, D. (1960). Credit risk and credit rationing. *Quarterly Journal of Economics , 74*, 258-278.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal , 7*, 281-312.

Hoggarth, G., Milne, A., & Wood, G. (1998). Alternative routes to banking stability:A comparison of UK and German banking systems. *Financial Stability Review , October*, 55-68.

Hondroyiannis, G., Lolos, S., & Papapetrou, E. (1999). Assessing competitive conditions in the Greek banking system. *Journal of International Financial Markets,Institutions and Money , 9*, 377–391.

Hughes, J., & Mester, L. (1993). A quality and risk-adjusted cost function for banks: evidence on the "too-big-to-fail" doctrine. *Journal of Productivity Analysis , 27*, 293-315.

Ikhide, S. I. (2000). *Efficiency of Commercial Banks in Namibia*. Retrieved June 2007, from Bank of Namibia:

http://www.bon.com.na/docs/pub/Efficiency%20of%20Commercial%20Banks%20in%20Namibia.pdf.

Ikhide, S. I. (2008). Measuring the operational efficiency of commercial banks in Namibia. *South African Journal of Economics , 76*, 586-595.

IMF and World Bank. (2002). *Financial Sector Assessment Programme: Selected Financial Sector Issues.* Washington D.C: International Monetary Fund/World Bank.

Isşık, I., & Akçaoğlu, E. (2006). An empirical analysis of productivity developments in "Traditional Banks": The initial post-liberalisation experience. *Central bank of the Republic of Turkey Review , 6*, 1-35.

Iwata, G. (1974). Measurement of conjectural variations in oligopoly. *Econometrica , 42*, 947-966.

Jackson, W. (1997). Market structure and the speed of adjustment: evidence of non-monotonicity. *Review of Industrial Organisation , 12*, 37-57.

Jackson, W. (1992). The price-concentration relationship in banking: a comment. *Review of Economics and Statistics , 74*, 373-376.

Jagtiani, J., & Khanthavit, A. (1996). Scale and scope economies at large banks: including off-balance sheet products and regulatory effects 1984-1991. *Journal of Banking and Finance , 20*, 1271-1287.

Jameson, K. (1972). Comment on the theory and measurement of dynamic X-efficiency. *The Quarterly Journal of Economics , 86*, 313-326.

Jayaratne, & Stratan. (1998). Entry restrictions, industry evolution, and dynamic efficiency: evidence from commercial banking. *Journal of Law and Economics, XLI, 239-273 , 41*, The Journal f Law and Economics.

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

Kablan, S. (2007). *Measuring banking efficiencyin developing countries: the case of the Western African Monetary Union (WAEMU).* Retrieved February 2008, from African Development Bank: http://www.afdb.org/pls/portal/docs/PAGE/ADB_ADMIN_PG/DOCUMENTS/AEC/MESURE%20DE%20LA%20PERFORMANCE%20DES%20BANQUES%20UEMOA.PDF

Kaparakis, E. I., Miller, S. M., & Noulas, A. G. (1994). Short-run cost inefficiency of commercial banks: a flexible stochastic frontier approach. *Journal of Money, Credit and Banking , 26*, 873-894.

Kasekende, L., Mlambo, K., Murinde, V., & Zhao, T. (2009). Restructuring for competitiveness-the financial services sector in Africa's largest economies. In W. E. Forum, W. Bank, & A. D. Bank, *African Competitiveness Report* (pp. 49-82). Geneva: World Economic Forum.

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *The American Economic Review , 80*, 1183-1200.

Kindleberger, C. (1978). *Mania, panics, and crashes: a history of financial crises.* New York: Basic Books.

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

Klein, M. (1971). A theory of the banking firm. *Journal of Money,Credit and Banking , 7*, 205-218.

Kodde, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica , 54*, 1243-1248.

Konings, J., Cayseele, V. P., & Warzynski, F. (2005). The effects of privatisation and competitive pressure on firms price-cost margins: micro evidence from emerging economies. *The Review of Economics and Statistics , 87* (1), 124-143.

Kubo, K. (2006). The degree of competition in the Thai banking industry before and after the East Asian crisis. *ASEAN Economic Bulletin , 23*, 325-340.

Kwan, S., & Eisenbeis, R. (1996). An analysis of inefficiency in banking: stochastic frontier approach. *Economic Review , 2*, 16-26.

Lau, L. (1982). On identifying the degree of competitiveness from industry price and output data. *Economics Letters , 10*, 93-99.

Leibenstein, H. (1966). Allocative Efficiency vs. "X-Efficiency". *The American Economic Review , 56*, 392-415.

Leightner, J. E., & Lovell, K. (1998). The impact of financial liberalisation of the performance of Thai banks. *Journal of Economics and Business , 50*, 115-131.

Leland, H., & Pyle, D. (1977). Informational asymmetries, financial structure and financial intermediation. *Journal of Finance , 32*, 371-387.

Lensink, R., Meesters, A., & Naaborg, I. (2008). Bank efficiency and foreign ownership: do good institutions matter? *Journal of Banking and Finance , 32*, 834–844.

Lerner, A. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies , 1*, 157-175.

Lim, G., & Randhawa, D. S. (2005). Competition, liberalisation and efficiency: evidence from a two-stage banking model on banks in Hong Kong and Singapore. *Managerial Finance , 31*, 52-77.

Lin, P.-w. (2005). An empirical analysis of bank mergers and cost efficiency in Taiwan. *Small Business Economics , 25*, 197-206.

Maimbo, M. S. (2002). The diagnosis and prediction of bank failures in Zambia,1990-1998. *Development Policy Review , 20*, 261-278.

Mamatzakis, E., Staikouras, C., & Koutsomanoli-Fillipak, N. (2005). Competition and concentration in the banking sector of the South Eastern European region. *Emerging Markets Review , 6*, 192-209.

Marcel, C., van Dijk, M., Lemmen, J., de Mooij, R., & Weigand, J. (2001, October). Competition and financial stability. *CBP Document, No. 15* . Netherlands Bureau for Economic Policy Anaysis.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance , 7*, 77-91.

Marshall, A. (1961). *Principles of Economics* (9 ed., Vol. 1). (C. Guillebaud, Ed.) London: Macmillan.

Matthews, K., Murinde, V., & Zhao, T. (2007). Competitive conditions among the major British banks. *Journal of Banking and Finance , 31*, 2025-2042.

Mckinnon, R. I. (1973). *Money and Capital in Economic Development.* Washington: The Brookings Institution).

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production function with composed error. *Internantional Economic Review , 18*, 435-444.

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

Miller, M., & Modigliani, F. (1961). Dividend policy, growth and the valuation of shares. *Journal of Business , 34*, 1-33.

Ministry of Finance and National Planning. (2004). *Financial Sector Development Plan for Zambia.* Lusaka: Government of the Republic of Zambia (GRZ).

Mktrtchyan, A. (2005). The evolution of competition in a transition economy: an application of the Panzar-Rosse model to Armenia. *The European Journal of Comparative Economics , 2*, 67-82.

Molyneux, P., Lloyd-Willians, D., & Thornton, J. (1994). Competitive conditions in European banking. *Journal of Banking and Finance , 18*, 445–459.

Monti, M. (1972). Deposit, credit and interest rate determination under alternative bank objective functions. In K. Shell, & P. G. Szego (Eds.), *Mathematical Methods in Invesment and Finance* (pp. 431-454). Amsterdam: North Holland.

Moore, W., & Craigwell, R. (2002). Market power and interest rate spreads in the Caribbean. *International Review of Applied Economics , 16*, 391-405.

Moshe, K., Kristiansen, E. G., & Bent, V. (2005). What determines banks' market power: Akerlof versus Herfindahl. *Working Paper, ANO 2005/8* . Oslo: Research Department, Norges Bank.

Mugume, A. (2007). Market structure and performance in Uganda's banking industry. *Paper presented at the African Econometrics Society* . Cape Town.

Murphy, N. B. (1972). Costs of banking activities: interactions between risk and operating costs, a comment. *Journal of Money, Credit and Banking , 4*, 614-615.

Mutoti, N. (2006). Monetary Policy Transmission in Zambia. *Bnak of Zambia Working Paper, No. WP/06/2006* . Bank of Zambia.

Mwilwa, C. N. (2007). Interest rate spreads in a liberalised environment: an empirical assessment on Zambia. *Unpublished Masters Degree Dissertation* . Centre of Development Studies,University of Glasgow.

Nannyonjo, J. (2002). Financial Sector reforms in Uganda (1990-2000): Interest rate spreads, market structure, bank performance and monetary policy. *Unpublished PhD Dissertation* . Stockholm: Economics Department, University of Goteborg.

Nathan, A., & Neave, E. H. (1989). Competition and contestability in Canada's financial system: empirical results. *Canadian Journal of Economics , 22*, 576–594.

Neumark, D., & Sharpe, S. (1992). Market structure and the nature of price rigidity: evidence from the market for consumer deposits. *Quarterly Journal of Economics , 107*, 657-680.

Ngugi, W. R. (2001). An empirical analysis of interest rate spread in Kenya . *Research Paper No. 106* . African Economic Research Consortium (AERC).

Northcott, C. A. (2004). Competition in banking: a review of the literature. *Working Paper No. 04-24* . Bank of Canada.

Okealaham, C. (2007). Estimating market power in the South African banking sector. *International Review of Applied Economics , 21*, 665-689.

Okealaham, C. (2004). The banking sector in Southern Africa: a test of the structure-performance hypothesis. *International Journal of Management and Decision Making , 5*, 216-245.

Oliver, M. A., Fumas, S. V., & Saurina, J. (2006). Risk premium and market power in credit markets. *Economics Letters , 93*, 450-456.

Oxford Policy Management. (2007). *FinMark Trust: Supply side study of the inclusiveness of Zambia's Financial System.* Lusaka: Government of Zambia (GRZ).

Pagano, M. (1993). Financial markets and growth: an overview. *European Economic Review , 37,* 613-622.

Panzar, J., & Rosse, J. (1982). Structure, conduct and comparative statics. *Bell Laboratories Economic Discussion Paper* . Bell Laboratories.

Panzar, J., & Rosse, J. (1987). Testing for "Monopoly" equilibrium. *The Journal of Industrial Economics , 35*, 443-456.

Peel, D. (1974). A note of X-Inefficiency. *The Quarterly Journal of Economics , 88*, 687-688.

Peltzman, S. (1977). The gains and losses from industrial concentration. *Journal of Law and Economics , 20*, 229-264.

Perera, S., Skully, M., & Wickramanayake, J. (2006). Competition and structure of South Asian banking: a revenue behaviour approach. *Applied Financial Economics , 16*, 789-801.

Pesek, B. (1970). Bank's supply function and the equilibrium quantity of money. *The Canadian Journal of Economics , 3*, 357-385.

Prince, M. Y., & Thurik, A. R. (1995). Do small firms' price-cost margins follow those of large firms? *Bulletin of Economic Research , 47*, 321-327.

Pringle, J. J. (1974). The imperfect-markets model of commercial bank financial management. *The Journal of Financial and Quantitative Analysis , 9*, 69-87.

Prisman, E. Z., Slovin, M., & Sushka, M. (1986). A general model of the banking firm under conditions of monopoly,uncertainty and recourse. *Journal of Monetary Economics , 17*, 293-304.

Pruteanu-Podpiera, A., Weill, L., & Schobert, F. (2008). Banking competition and cost efficiency: A Micro-Data Analysis on the Czech banking industry. *Comparative Economic Studies , 50*, 253-273.

Pyle, D. H. (1971). On the theory of financial intermediation. *Journal of Finance , 26*, 737-747.

Rangan, N., Grabowski, R., Aly, H., & Pasurka, C. (1988). The technical efficiency of U.S. banks. *Economics Letters , 28*, 169-175.

Rao, A. (2005). Cost frontier efficiency and risk-return analysis in an emerging market. *International Review of Financial Analysis , 14*, 283-303.

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.

Rhoades, S. A. (1992). Evidence on the size of banking markets from mortgage loan rates in twenty cities. *Federal Reserve Bulletin , 78*, 117-118.

Rochet, J. (1992). Capital requirements and the behaviour of commercial banks. *European Economic Review , 36* (5), 1137-1178.

Rosse, J., & Panzar, J. C. (1977). Chamberlin vs. Robinson: An empirical test for monopoly rents. *Bell Laboratories Economic discussion 90* . Bell Laboratories.

Saab, S. Y., & Vacher, J. (2007). Banking Sector Integration and Competition in CEMAC. *IMF Working Paper WP/07/3* . International Monetary Fund.

Samad, A. (2008). Market structure, conduct and performance: evidence from the Bangladesh banking industry. *Journal of Asian Economics , 19*, 181–193.

Santomero, F. (1984). Modelling the banking firm: a survey. *Journal of Money Credit and Banking , 16*, 576-602.

Santos, J. A. (2000). Bank capital regulation in contemporary banking theory: a review of the literature. *BIS Working Papers, No. 90* . Bank for International Settlements (BIS).

Sarkar, J., & Bhaumik, S. K. (1998). Deregulation and the limits to banking market competition: some insights from India. *International Journal of Development Banking , 16*, 29-42.

Sarkar, J., Sarkar, S., & Bhaumik, S. (1998). Does ownership always matter: evidence from the Indian banking industry. *Journal of Comparative Economics , 26*, 262-281.

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

Seelanatha, S. L. (2007). Efficiency, productivity change and market structure of the banking industry in Sri Lanka. *Unpublished PhD Dissertation* . University of Southern Queensland, Australia.

Sensarma, R. (2006). Are foreign banks always the best? Comparison of state-owned, private and foreign banks in India. *Economic Modelling , 23*, 717-735.

Shaffer, S. (1993). A test of competition in Canadian banking. *Journal of Money, Credit and Banking , 25*, 49-61.

Shaffer, S. (1994). Banking competition in concentrated markets. *Business Review , March*, 3-16.

Shaffer, S. (2001). Banking conduct before the European single banking licence: a cross country comparison. *North American Journal of Economics and Finance , 12*, 79-104.

Shaffer, S. (1983). Non-structural measures of competition: toward a synthesis of alternatives. *Economics Letters , 12*, 349-353.

Shanmugan, K., & Das, A. (2004). Efficiency of Indian banks during the reform period. *Applied Financial Economics , 14*, 681-686.

Sharpe, W. (1963). A simplified model for portfolio analysis. *Management Science , 2*, 277-293.

Shaw, E. (1973). *Financial Deepening in Economic Development.* New York: Oxford University Press.

Sherman, D., & Gold, F. (1985). Bank branch operating efficiency: evaluation with Data Envelopment Analysis. *Journal of Banking and Finance , 9*, 297-315.

Sherman, H., & Gold, F. (1985). Bank branch operating efficiency: Evaluation with data envelopment analysis. *Journal of Banking and Finance , 9*, 297-315.

Sjöberg, P. (2006). Two essays on performance and growth in Swedish banking. *Unpublished PhD Dissertation* . Department of Economics,Göteborg University.

Small, I. (1998). The cyclicality of mark-ups and profit margins: some evidence for manufacturing and services. *Quarterly Bulletin* (August).

Solis, L., & Maudos, J. (2008). The social costs of bank marker power: evidence from Mexico. *Journal of Comparative Economics , 36*, 467-488.

Staikouras, C., & Koutsomanoli-Filipaki, A. (2006). Competition and concentration in the new European banking landscape. *European Financial Management , 12*, 443-482.

Staikouras, C., & Wood, G. (2000). Competition and banking stability in the Euro area:The cases for Greece and Spain. *The Journal of International Banking Regulation , 2*, 7-24.

Staikouras, C., Mamatzakis, E., & Koutsomanoli-Filippaki, A. (2008). Cost efficiency of the banking industry in the South Eastern European region. *Journal of International Financial Markets, Institutions and Money (In Press, Corrected Proof) , 18*, 483-497.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics , 13*, 57-66.

Stigler, G. (1976). "The Xistence of X-Efficiency". *American Economic Review , 66*, 213-216.

Stiglitz, J. (1994). "The role of the State in financial markets". *Proceedings of the World Bank Annual Conference of Development Economics* (Supplement to the World Bank Economic Review and World Bank Research Observer), pp. 19-52.

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

Sturma, J.-E., & Williams, B. (2004). Foreign bank entry, deregulation and bank efficiency: lessons from the Australian experience. *Journal of Banking & Finance , 28*, 1775-1799.

Swank, J. (1996). Theories of the banking firm: a review of the literature. *Bulletin of Economic Research , 48*, 173-207.

Telser, L. G. (2007). Solvency versus competition: Hobson's choice for the Fed. *Journal of International Money and Finance , 26*, 1151-1773.

Thakor, A. V. (1996). The design of financial systems: an overview. *Journal of Banking and Finance , 20*, 917-948.

Tirole, J. (1992). *The Theory of Industrial Organisation.* Cambridge, MA: MIT Press.

Tobin, J. (1958). Liquidity preference as behaviour towards risk. *Review of Economic Studies , 26*, 65-86.

Toolsema, L. (2004). Monetary policy and market power in banking. *Journal of Economics , 83*, 71-83.

Toolsema, L. (2003). On competition and banking. *Unpublished PhD dissertation* . University of Gronigen.

Tovar, J. A., Jaramillo, C. R., & Hernandez, C. E. (2007). Risk, concentration and market power in the banking industry: evidence from the Colombian System (1997-2006). *Serie Documentos Cede, 2007-27* . Centro de Estudios sobre Desarrollo Económico.

Uiboupin, J. (2004). Effects of foreign banks entry on bank performance in the CEE countries. *Working Paper No.33* . University of Tartu - Faculty of Economics and Business Administration.

van Damme, E. (1994). Banking: A survey of recent microeconomic theory. *Oxford Review of Economic Policy , 10*, 14-20.

VanHoose, D. (1988). Deregulation and oligopolistic rivalry in bank deposit markets. *Journal of Banking and Finance , 12*, 379-388.

Vera, L., Zambrano-Sequin, L., & Faust, A. (2007). The efficiency-stability trade-off:the case of high interest rate spreads in Venezuela. *The Developing Economies , XLV*, 1-26.

Vesala, J. (1995). Testing for competition in banking: behavioral evidence from Finland. *Bank of Finland Studies, E.1* . Bank of Finland.

Weill, L. (2003). Banking efficiency in transition economies: the role of foreign ownership. *The Economics of Transition , 11*, 569-592.

Weill, L. (2007). Is there gap in bank efficiency between CEE and Western European countries? *Comparative Economic Studies , 49*, 101-127.

Weill, L. (2004). Measuring cost efficiency in European banking: A comparison of frontier techniques. *Journal of Productivity Analysis , 21*, 133-152.

Wong, J., Wong, E., Fong, T., & Choi, K. F. (2007). Testing for collusion in the Hong Kong banking sector. *Research Memorandum, 03/2007* . Hong Kong Monetary Authority (HKMA).

World Bank. (2006). Review of World Bank Assistance for Financial Sector Reforms. Washington D.C: Independent Evaluation Group, World Bank.

World Bank. (2004). *Zambia-Country Economic Memorandum: Policies for Growth and Diversification.* Washington DC: World Bank.

Worthington, A. (1998). Efficiency in Australian building societies:an econometric cost function approach using panel data. *Applied Financial Economics , 8*, 459-467.

Yao, S., Jiang, C., Feng, G., & Willenbockel, D. (2007). WTO challenges and efficiency of Chinese banks. *Applied Economics , 39*, 629-643.

Yildirim, C. (2002). Evolution of banking efficiency within an unstable macroeconomic environment: the case of Turkish commercial banks. *Applied Economics , 34*, 2289-2301.

Yildirim, S. H., & Philippatos, G. C. (2007). Competition and contestability in central and eastern European banking markets. *Managerial Finance , 33*, 195-209.

Yildirim, S. H., & Philippatos, G. (2007). Efficiency of banks: recent evidence from the transition economies of Europe 1993-2000. *The European Journal of Finance , 13*, 123-143.

Yuan, Y. (2006). The state of competition of the Chinese banking industry. *Journal of Asian Economics , 17*, 519-534.

Zaim, O. (1995). The effect of financial liberalisation on the efficiency of Turkish commercial banks. *Applied Financial Economics , 5*, 257-264.

Zajc, P. (2006). A comparative study of bank efficiency in central and eastern Europe: The role of foreign ownership. *International Finance Review , 6*, 117-156.

Zardkoohi, A., Nanda, R., & James, K. (1986). Homogeneity restrictions on the translog cost model: a note. *The Journal of Finance , XLI*, 1153-1155.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association , 57*, 348-368.

Appendices

Appendix I: Selected policy reforms: 1992-2003

Liberalised of commercial banks' interest rates	Sept. 1992
Exchange rates unified	Dec. 1992
Introduction of Open Market Operations in primary Treasury bills	Jan. 1993
28-day Treasury bills auctions introduced	Mar. 1993
182 day Treasury bills introduced	Aug. 1993
Bank of Zambia foreign exchange dealing system introduced	Dec. 1993
BoZ rate on unauthorised overdrafts increased by 5 % to 20 % above the 28-day yield rate	Dec. 1993
Shortfall on statutory reserves in excess of K1.0 billion set at 70 % above the 28-day yield rate	Dec. 1993
Foreign Exchange Controls Act revoked	Jan. 1994
Statutory reserve requirements for foreign currency accounts equal to domestic currency accounts	Feb. 1994
The Banking and Financial Services Act (BFSA) version of 1995 introduced	Jun. 1994
Government bond auctions introduced	Dec. 1994
Introduction of daily BoZ Open Market Operations in credit and deposit auctions	Mar. 1995
Abolition of Off-Tender window and commercial banks become sole participants at Treasury bill auctions	Dec. 1996
Enactment of the Bank of Zambia Act No. 43	Dec. 1996
Surplus funds from the 1996 fiscal budget set aside to liquidate Treasury bills stock to reduce interest rates	Jan. 1997
Government bonds listed on the Lusaka stock Exchange	Mar. 1998
18-month Government bonds re-introduced	Oct. 1998
Risk-based supervision implemented	1999
Introduction of 273-day Treasury bills and 24 months Government bond	Oct. 2000
Early warning system introduced	2000
BoZ introduced new guidelines for foreign exchange access and dealing	Feb. 2001
Conversion of kwacha equivalent foreign currency reserve balances into foreign currency reserve deposits commenced	Apr. 2001
Broad based foreign exchange interbank system introduced	Jul. 2003

Source: Bank of Zambia

Appendix I-1a: Total revenue PR results for domestic and foreign banks

	Dependent Variable: $\ln(TREVASST)$					
	Domestic banks			Foreign banks		
Variable	Coefficient	t-statistic	p-value	Coefficient	t-statistic	p-value
Intercept	-1.936	-1.191	0.235	-2.105	-2.718	0.005***
Input prices						
$\ln(w_L)$	0.255	2.854	0.005***	0.307	4.277	0.000***
$\ln(w_F)$	0.194	4.465	0.000***	0.293	9.418	0.000***
$\ln(w_K)$	0.143	2.957	0.004***	0.084	1.793	0.075*
Bank-specific and regulatory factors						
ln(*OPPCOST*)	-0.066	-1.036	0.301	-0.089	-3.051	0.003***
ln(*CAPRATIO*)	0.001	0.039	0.969	-0.019	-0.407	0.685
ln(*RISK*)	0.004	-0.158	0.874	-0.041	-1.828	0.069*
ln(*INTERMED*)	-0.008	-0.162	0.872	0.073	1.219	0.225
ln(*BRANCH*)	-0.062	-0.418	0.676	0.059	0.696	0.488
Macroeconomic variables						
ln(*INFLATION*)	-0.125	-1.659	0.099*	-0.191	-2.834	0.005***
ln(*EXR*)	0.166	1.080	0.282	0.242	2.564	0.011***
ln(*TBR*)	0.121	1.246	0.215	0.137	2.814	0.005***
H-statistic	0.591			0.684		
H=0 (p-value)	0.000***			0.000***		
H=1 (p-value)	0.000***			0.001***		
Market Structure	Monopolistic Competition			Monopolistic Competition		
Diagnostics						
Hausman Test (p-value)	0.000***			0.000***		
σ_u	0.248			0.348		
σ_ε	0.216			0.172		
ρ	0.528			0.804		
R^2	0.580			0.472		
No. of Obs	195			184		
F-statistic (p-value)	0.000***			0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent

$\ln(TREVASST)$ - Log of Total Interest Revenue/Total Assets

Source: Bank of Zambia and author's own estimations

Appendix I-1b: Total revenue PR results for domestic and foreign banks

	Dependent Variable: $\ln\left(TREVASST\right)$					
	Small banks			Large banks		
Variable	Coefficient	t-statistic	p-value	Coefficient	t-statistic	p-value
Intercept	-1.905	-1.991	0.048**	-3.088	-1.717	0.089*
Input prices						
$\ln(w_L)$	0.309	4.932	0.000***	0.183	1.782	0.078*
$\ln(w_F)$	0.276	8.774	0.000***	0.153	3.115	0.002***
$\ln(w_K)$	0.083	2.198	0.029**	0.196	2.399	0.018**
Bank-specific and regulatory factors						
ln(*OPPCOST*)	-0.064	-1.999	0.047**	-0.199	-1.821	0.072*
ln(*CAPRATIO*)	0.003	0.078	0.938	0.082	1.153	0.252
ln(*RISK*)	-0.016	-0.839	0.402	-0.009	-0.221	0.825
ln(*INTERMED*)	0.076	1.677	0.095*	-0.075	-0.939	0.350
ln(*BRANCH*)	0.035	0.031	0.726	0.129	1.118	0.267
Macroeconomic variables						
ln(*INFLATION*)	-0.171	-2.681	0.008***	-0.196	-1.781	0.078*
ln(*EXR*)	0.204	1.864	0.064*	0.300	1.233	0.221
ln(*TBR*)	0.095	1.843	0.067*	0.295	2.044	0.044**
H-statistic	0.668			0.533		
H=0 (p-value)	0.000***			0.000***		
H=1 (p-value)	0.000***			0.001***		
Market Structure	Monopolistic Competition			Monopolistic Competition		
Diagnostics						
Hausman Test (p-value)	0.099*			0.079*		
σ_u	0.280			0.075		
σ_ε	0.198			0.181		
ρ	0.667			0.146		
R^2	0.655			0.645		
No. of Obs.	270			109		
F-statistic (p-value)	0.000***			0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent

$\ln\left(TREVASST\right)$ - Log of Total Interest Revenue/Total Assets

Source: Author's own estimations based on Bank of Zambia data

Appendix I-2a: Interest revenue PR results for domestic and foreign banks

	Dependent Variable: $\ln(INTRASST)$					
	Domestic banks			Foreign banks		
Variable	Coefficient	t-statistic	p-value	Coefficient	t-statistic	p-value
Intercept	-6.117	-5.183	0.000***	-4.183	-5.983	0000***
Input prices						
$\ln(w_L)$	0.350	4.448	0.000***	0.175	2.914	0.004***
$\ln(w_F)$	0.198	3.604	0.000***	0.379	12.211	0.000***
$\ln(w_K)$	0.135	2.253	0.026**	0.024	0.523	0.601
Bank-specific and regulatory factors						
ln(*OPPCOST*)	-0.124	-2.259	0.025**	-0.071	-2.487	0.014***
ln(*CAPRATIO*)	-0.026	-0.569	0.570	-0.080	-1.808	0.072*
ln(*RISK*)	-0.046	-2.032	0.044**	-0.022	-0.968	0.334
ln(*INTERMED*)	0.006	0.095	0.925	0.124	2.205	0.029**
ln(*BRANCH*)	0.277	1.814	0.071*	0.112	1.335	0.184
Macroeconomic variables						
ln(*INFLATION*)	-0.184	-2.008	0.046**	-0.135	-1.956	0.052**
ln(*EXR*)	0.578	4.526	0.000***	0.345	3.559	0.000***
ln(*TBR*)	0.233	2.618	0.010***	0.107	2.288	0.023**
H-statistic	0.683			0.577		
H=0 (p-value)	0.000***			0.000***		
H=1 (p-value)	0.004***			0.000***		
Market Structure	Monopolistic Competition			Monopolistic Competition		
Diagnostics						
Hausman Test (p-value)	0.047**			0.028***		
σ_u	0.423			0.344		
σ_ε	0.240			0.162		
ρ	0.756			0.819		
R^2	0.347			0.540		
No. of Obs.	196			184		
F-statistic (p-value)	0.000***			0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent

$\ln(INTRASST)$ - Log of Total Interest Revenue/Total Assets

Source: Bank of Zambia and author’s estimations

Appendix I-2b: Interest revenue PR results for small and large banks

	Dependent Variable: $\ln\left(INTRASST\right)$					
	Small banks			Large banks		
Variable	Coefficient	t-statistic	p-value	Coefficient	t-statistic	p-value
Intercept	-5.178	-6.711	0.000***	-5.873	-3.519	0.001***
Input prices						
$\ln(w_L)$	0.268	4.755	0.000***	0.273	2.898	0.005***
$\ln(w_F)$	0.329	9.978	0.000***	0.162	3.175	0.002***
$\ln(w_K)$	0.089	1.984	0.048**	0.176	1.578	0.118
Bank-specific and regulatory factors						
ln(*OPPCOST*)	-0.077	-2.342	0.020**	-0.265	-2.604	0.011***
ln(*CAPRATIO*)	-0.038	-1.000	0.318	0.117	1.377	0.172
ln(*RISK*)	-0.063	-3.604	0.000***	-0.265	-2.604	0.011***
ln(*INTERMED*)	0.127	3.623	0.000***	-0.095	-0.911	0.365
ln(*BRANCH*)	0.218	2.337	0.020**	0.156	1.266	0.209
Macroeconomic variables						
ln(*INFLATION*)	-0.131	-1.812	0.071*	-0.384	-2.658	0.009***
ln(*EXR*)	0.461	4.277	0.000***	0.689	2.858	0.005***
ln(*TBR*)	0.150	2.779	0.006***	0.479	3.371	0.001***
H-statistic	0.686			0.611		
H=0 (p-value)	0.000***			0.000***		
H=1 (p-value)	0.000***			0.013***		
Market Structure	Monopolistic Competition			Monopolistic Competition		
Diagnostics						
Hausman Test (p-value)	0.000***			0.093*		
σ_u	0.311			0.128		
σ_ε	0.204			0.215		
ρ	0.699			0.262		
R^2	0.507			0.532		
No. of Obs.	271			109		
F-statistic (p-value)	0.000***			0.000***		

Significance level: *** 1 percent, ** 5 percent and * 10 percent

$\ln\left(INTRASST\right)$ - Log of Total Interest Revenue/Total Assets

Source: Bank of Zambia and author’s estimations

Appendix III: Cost efficiency scores, quarterly estimates

Quarter	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12	Bank 13	Bank 14	Bank 15	(All Banks)
1998q1	0.917	0.943	0.943	…	…	-	0.914	…	0.759	0.860	0.935	0.534	0.379	0.910	-	0.809
1998q2	0.950	0.963	0.949	…	…	-	0.896	0.799	0.905	0.888	…	0.916	0.373	0.934	-	0.857
1998q3	0.890	0.965	0.940	…	…	-	0.786	…	0.961	0.707	0.966	0.952	0.409	0.965	-	0.854
1998q4	0.943	0.941	0.964	…	…	-	0.883	…	0.920	0.873	0.951	0.943	0.330	…	-	0.861
1999q1	0.937	0.964	0.967	…	…	-	0.949	…	0.954	0.929	0.957	0.945	…	0.923	-	0.947
1999q2	0.935	0.975	0.959	…	…	-	0.949	…	0.956	0.887	0.957	0.972	0.617	0.898	-	0.910
1999q3	0.951	0.969	0.961	…	…	-	0.949	…	0.964	0.892	0.949	0.974	0.684	0.911	-	0.921
1999q4	0.961	0.959	0.976	…	…	-	0.944	…	0.961	0.957	0.943	0.956	0.678	0.911	-	0.925
2000q1	0.971	0.976	0.961	…	…	…	0.959	…	…	0.937	0.948	0.970	0.624	0.945	-	0.921
2000q2	0.972	0.974	0.946	…	…	…	0.896	…	…	0.964	0.915	0.960	0.650	0.955	-	0.915
2000q3	0.977	0.977	0.947	…	…	…	0.927	…	0.946	0.957	0.934	0.922	0.579	0.953	-	0.912
2000q4	0.977	0.937	0.972	…	…	…	0.946	…	0.890	0.898	0.945	0.934	0.506	0.901	-	0.890
2001q1	0.960	0.942	0.969	…	…	…	0.936	…	0.960	0.922	0.963	0.959	0.748	0.957	-	0.932
2001q2	0.958	0.953	0.970	…	…	…	0.938	…	0.951	0.945	0.975	0.967	0.751	0.963	-	0.937
2001q3	0.952	0.977	0.975	…	…	…	0.926	…	…	0.941	0.927	0.970	0.702	0.968	-	0.927
2001q4	0.947	0.957	0.979	…	…	0.979	0.937	…	…	0.980	0.948	0.932	0.936	0.939	-	0.953
2002q1	0.961	0.964	0.978	…	…	0.945	0.894	…	…	0.977	0.954	0.920	0.941	0.966	-	0.950
2002q2	0.900	0.904	0.977	…	…	0.739	0.945	…	0.966	0.980	0.953	0.884	0.857	0.965	-	0.915
2002q3	0.871	0.944	0.972	…	…	0.762	0.969	…	0.951	0.978	0.960	0.916	0.797	0.932	-	0.914
2002q4	0.841	0.959	0.973	…	…	0.817	0.928	…	0.931	0.974	0.822	0.943	0.834	0.918	-	0.904
2003q1	0.952	0.953	0.975	…	…	0.723	0.860	…	0.966	0.949	0.976	0.941	0.962	0.960	-	0.929
2003q2	0.955	0.936	0.952	…	…	0.775	0.971	…	0.962	0.966	0.958	0.937	0.944	0.966	-	0.938
2003q3	0.940	0.942	0.958	…	…	0.930	0.972	…	0.964	0.968	0.977	0.936	0.941	0.966	-	0.954
2003q4	0.884	0.956	0.957	…	…	0.538	0.952	…	0.963	0.983	0.967	0.913	0.916	0.950	-	0.907

Notes:

… Missing data for some variables resulting in no estimated scale economies

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix III-1 *continued*

2004q1	0.954	0.962	0.954	…	0.736	0.734	0.773	…	0.962	0.957	0.917	0.885	-	-	0.931	0.888
2004q2	0.939	0.947	0.921	…	0.703	0.753	0.871	…	0.933	0.950	0.942	0.848	-	-	…	0.881
2004q3	0.939	0.964	0.974	…	0.738	0.668	0.883	…	0.966	0.954	0.964	0.948	-	-	0.949	0.904
2004q4	0.912	0.970	0.967	…	0.803	0.686	0.883	…	0.973	0.950	0.982	0.964	-	-	0.964	0.914
2005q1	0.902	0.970	0.973	…	0.705	0.928	0.906	…	0.974	0.958	0.979	0.971	-	-	0.968	0.930
2005q2	0.688	0.971	0.970	…	0.693	0.936	0.941	…	0.963	0.964	0.976	0.957	-	-	0.941	0.909
2005q3	0.933	0.967	0.970	0.972	0.854	0.925	0.921	…	0.968	0.952	0.980	0.949	-	-	0.945	0.945
2005q4	0.951	0.965	0.966	0.967	0.782	0.902	0.950	…	0.956	0.934	0.942	0.969	-	-	0.956	0.937
2006q1	0.964	0.976	0.976	0.974	0.925	0.925	0.947	…	0.959	0.968	0.980	0.979	-	-	0.969	0.962
2006q2	0.962	0.975	0.957	0.951	0.900	0.952	0.968	…	0.948	0.955	0.981	0.977	-	-	…	0.957
2006q3	0.963	0.981	0.948	0.934	0.893	0.946	0.969	…	0.957	0.956	0.984	0.973	-	-	0.970	0.956
2006q4	0.901	0.982	0.923	0.974	0.934	0.931	0.975	…	0.957	0.964	0.979	0.974	-	-	0.964	0.955
Mean	0.931	0.960	0.962	0.962	0.805	0.833	0.923	0.799	0.947	0.938	0.954	0.933	-0.008	-0.006	0.956	0.917

Notes:

… Missing data for some variables resulting in no scale economies estimates

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix IV-1: Economies of scale, quarterly estimates

Quarter	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12	Bank 13	Bank 14	Bank 15	Mean Scale Economies (All `Banks)
1998q1	1.114	1.127	1.149	…	…	-	1.214	…	1.085	1.182	1.242	1.316	1.256	1.336	-	1.202
1998q2	1.102	1.114	1.150	…	…	-	1.225	1.337	1.076	1.199	…	1.315	1.243	1.332	-	1.209
1998q3	1.094	1.092	1.132	…	…	-	1.200	…	1.083	1.178	1.243	1.297	1.232	…	-	1.172
1998q4	1.105	1.091	1.114	…	…	-	1.178	…	1.075	1.166	1.260	1.290	1.199	1.345	-	1.182
1999q1	1.086	1.074	1.106	…	…	-	1.158	…	1.067	1.163	1.233	1.297	…	1.320	-	1.167
1999q2	1.070	1.068	1.086	…	…	-	1.151	…	1.059	1.167	1.237	1.290	1.271	1.321	-	1.172
1999q3	1.062	1.064	1.078	…	…	-	1.137	…	1.051	1.159	1.227	1.273	1.265	1.334	-	1.165
1999q4	1.047	1.076	1.087	…	…	-	1.156	…	1.054	1.156	1.224	1.303	1.239	1.328	-	1.167
2000q1	1.064	1.082	1.083	…	…	…	1.134	…	…	1.150	1.223	1.282	1.257	1.320	-	1.177
2000q2	1.056	1.080	1.084	…	…	…	1.145	…	…	1.146	1.203	1.279	1.254	1.308	-	1.173
2000q3	1.056	1.073	1.078	…	…	…	1.138	…	1.061	1.147	1.199	1.265	1.252	1.293	-	1.156
2000q4	1.061	1.051	1.081	…	…	…	1.137	…	1.045	1.143	1.198	1.254	1.239	1.301	-	1.151
2001q1	1.031	1.055	1.078	…	…	…	1.120	1.205	1.047	1.144	1.204	1.238	1.225	1.306	-	1.150
2001q2	1.024	1.044	1.065	…	…	…	1.095	1.182	1.038	1.132	1.175	1.222	1.217	1.301	-	1.136
2001q3	1.027	1.043	1.057	…	…	…	1.079	1.188	…	1.111	1.166	1.214	1.255	1.284	-	1.142
2001q4	1.028	1.044	1.056	…	…	1.184	1.086	1.222	…	1.101	1.141	1.214	1.266	1.284	-	1.148
2002q1	1.031	1.042	1.050	…	…	1.143	1.085	1.219	…	1.096	1.138	1.193	1.261	1.278	-	1.140
2002q2	1.031	1.040	1.046	…	…	1.136	1.071	1.217	1.024	1.095	1.137	1.199	1.272	1.271	-	1.128
2002q3	1.030	1.043	1.043	…	…	1.130	1.080	1.252	1.023	1.088	1.148	1.184	1.265	1.269	-	1.129
2002q4	1.022	1.026	1.046	…	…	1.102	1.059	1.246	1.018	1.079	1.162	1.172	1.251	1.262	-	1.120
2003q1	1.024	1.032	1.044	…	…	1.096	1.070	1.240	1.021	1.080	1.155	1.174	1.263	1.255	-	1.121
2003q2	1.029	1.027	1.032	…	…	1.100	1.063	1.254	1.023	1.081	1.155	1.178	1.255	1.256	-	1.121
2003q3	1.026	1.030	1.029	…	…	1.099	1.067	1.260	1.021	1.075	1.154	1.179	1.237	1.260	-	1.120
2003q4	1.032	1.027	1.036	…	…	1.115	1.068	1.267	1.028	1.073	1.165	1.170	1.343	1.168	-	1.124

Notes:

… Missing data for some variables resulting in no estimated scale economies

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix IV-1 *continued*

2004q1	1.034	1.038	1.042	1.099	1.174	1.117	1.084	1.263	1.038	1.076	1.173	1.168	-	-	…	1.109
2004q2	1.033	1.033	1.040	1.092	1.179	1.125	1.093	1.272	1.036	1.075	1.175	1.161	-	-	…	1.110
2004q3	1.031	1.031	1.034	1.091	1.185	1.125	1.088	1.257	1.038	1.070	1.147	1.146	-	-	1.172	1.109
2004q4	1.025	1.034	1.036	1.072	1.170	1.116	1.083	1.239	1.035	1.064	1.158	1.119	-		1.149	1.100
2005q1	1.025	1.029	1.029	1.058	1.159	1.112	1.083	1.242	1.037	1.048	1.147	1.107	-	-	1.140	1.093
2005q2	1.020	1.032	1.024	1.039	1.177	1.112	1.074	1.253	1.031	1.045	1.147	1.102	-	-	1.132	1.091
2005q3	0.999	1.035	1.013	1.052	1.166	1.113	1.078	1.246	1.034	1.051	1.156	1.107	-	-	1.140	1.091
2005q4	1.009	1.019	1.024	1.045	1.153	1.117	1.080	1.215	1.034	1.051	1.147	1.102	-	-	1.125	1.086
2006q1	1.002	1.019	1.026	1.052	1.153	1.113	1.071	1.167	1.028	1.053	1.142	1.097	-	-	…	1.077
2006q2	1.004	1.016	1.023	1.036	1.143	1.116	1.070	1.125	1.030	1.035	1.130	1.091	-	-	…	1.068
2006q3	1.002	1.016	1.009	1.002	1.137	1.113	1.061	1.114	1.028	1.029	1.120	1.090	-	-	1.123	1.065
2006q4	0.998	1.151	1.153	1.366	1.119	1.211	1.358	1.089	1.180	1.261	1.330	1.316	-	-	…	1.211
Mean	1.125	1.053	1.063	1.084	1.160	1.124	1.115	1.223	1.047	1.110	1.182	1.206	1.253	1.293	1.140	1.136

Notes:

… Missing data for some variables resulting in no scale economies estimates

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix IV-2: Technological progress, quarterly estimates

Quarter	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12	Bank 13	Bank 14	Bank 15	All Banks
1998q1	-0.011	-0.009	-0.011	…	…	-	-0.013	…	-0.011	-0.016	-0.017	-0.017	-0.016	-0.013	-	-0.013
1998q2	-0.011	-0.009	-0.006	…	…	-	-0.008	0.991	-0.013	-0.011	…	-0.017	-0.017	-0.014	-	0.089
1998q3	-0.010	-0.009	-0.008	…	…	-	-0.010	…	-0.009	-0.013	-0.016	-0.015	-0.016	…	-	-0.012
1998q4	-0.009	-0.009	-0.009	…	…	-	-0.013	…	-0.010	-0.012	-0.012	-0.013	-0.016	…	-	-0.011
1999q1	-0.009	-0.009	-0.010	…	…	-	-0.013	…	-0.011	-0.011	-0.013	-0.014	…	-0.012	-	-0.011
1999q2	-0.010	-0.009	-0.012	…	…	-	-0.013	…	-0.011	-0.011	-0.015	-0.016	-0.011	-0.011	-	-0.012
1999q3	-0.010	-0.009	-0.012	…	…	-	-0.013	…	-0.008	-0.011	-0.014	-0.016	-0.011	-0.009	-	-0.011
1999q4	-0.012	-0.006	-0.009	…	…	-	-0.010	…	-0.008	-0.010	-0.012	-0.012	-0.012	-0.008	-	-0.010
2000q1	-0.008	-0.005	-0.008	…	…	…	-0.010	…	…	-0.011	-0.010	-0.013	-0.011	-0.008	-	-0.009
2000q2	-0.006	-0.003	-0.007	…	…	…	-0.009	…	…	-0.009	-0.011	-0.011	-0.010	-0.007	-	-0.008
2000q3	-0.004	-0.003	-0.006	…	…	…	-0.008	…	-0.004	-0.007	-0.011	-0.011	-0.008	-0.007	-	-0.007
2000q4	-0.003	-0.004	-0.005	…	…	…	-0.008	…	-0.007	-0.007	-0.010	-0.011	-0.011	-0.008	-	-0.007
2001q1	-0.006	-0.003	-0.005	…	…	…	-0.008	-0.001	-0.006	-0.006	-0.011	-0.011	-0.012	-0.006	-	-0.007
2001q2	-0.007	-0.004	-0.005	…	…	…	-0.008	0.003	-0.006	-0.004	-0.009	-0.010	-0.010	-0.006	-	-0.006
2001q3	-0.005	-0.003	-0.005	…	…	…	-0.010	0.005	…	-0.007	-0.009	-0.009	-0.005	-0.005	-	-0.005
2001q4	-0.004	-0.002	-0.004	…	…	-0.010	-0.007	0.009	…	-0.005	-0.012	-0.009	-0.003	-0.006	-	-0.005
2002q1	-0.003	-0.001	-0.003	…	…	-0.010	-0.007	0.012	…	-0.004	-0.012	-0.009	-0.004	-0.005	-	-0.004
2002q2	-0.002	-0.002	-0.003	…	…	-0.007	-0.009	0.006	-0.003	-0.003	-0.011	-0.008	-0.003	-0.004	-	-0.004
2002q3	-0.001	-0.001	-0.003	…	…	-0.004	-0.005	0.013	-0.002	-0.001	-0.007	-0.009	-0.001	-0.003	-	-0.002
2002q4	-0.001	-0.001	-0.001	…	…	-0.009	-0.004	0.014	-0.003	-0.003	-0.005	-0.008	-0.001	-0.002	-	-0.002
2003q1	0.000	0.001	0.000	…	…	-0.007	-0.004	0.009	-0.001	-0.002	-0.003	-0.008	-0.001	-0.002	-	-0.002
2003q2	0.002	0.001	-0.001	…	…	-0.004	-0.003	0.014	-0.001	0.000	-0.005	-0.008	0.000	-0.002	-	-0.001
2003q3	0.002	0.002	0.000	…	…	-0.002	-0.002	0.015	0.001	0.003	-0.002	-0.006	0.003	0.000	-	0.001
2003q4	0.003	0.003	0.001	…	…	0.000	0.002	0.017	0.003	0.002	0.001	-0.004	-0.011	0.002	-	0.002

Notes:

… Missing data for some variables resulting in no estimated scale economies

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix IV-2 *continued*

2004q1	0.003	0.007	0.006	0.010	-0.004	-0.001	0.002	0.018	0.006	0.005	0.002	-0.002	-	-	…	0.004
2004q2	0.004	0.008	0.005	0.010	-0.004	0.002	0.003	0.019	0.007	0.005	0.002	0.000	-	-	…	0.005
2004q3	0.006	0.008	0.006	0.013	-0.003	0.004	0.005	0.016	0.009	0.004	0.000	0.001	-	-	0.006	0.006
2004q4	0.007	0.009	0.006	0.012	-0.003	0.003	0.004	0.015	0.007	0.005	0.002	0.001	-	-	0.005	0.006
2005q1	0.007	0.008	0.006	0.009	-0.001	0.002	0.004	0.017	0.008	0.004	0.003	0.001	-	-	0.006	0.006
2005q2	0.010	0.009	0.007	0.007	0.002	0.003	0.004	0.017	0.008	0.004	0.005	0.002	-	-	0.007	0.007
2005q3	0.006	0.010	0.005	0.009	0.004	0.003	0.006	0.016	0.010	0.006	0.009	0.004	-	-	0.010	0.008
2005q4	0.007	0.009	0.008	0.008	0.003	0.004	0.006	0.016	0.009	0.006	0.006	0.004	-	-	0.008	0.007
2006q1	0.006	0.010	0.008	0.008	0.004	0.004	0.006	0.014	0.009	0.007	0.006	0.005	-	-	…	0.007
2006q2	0.007	0.010	0.009	0.009	0.004	0.007	0.008	0.015	0.010	0.006	0.006	0.006	-	-	…	0.008
2006q3	0.009	0.012	0.009	0.008	0.005	0.008	0.009	0.017	0.011	0.008	0.009	0.007	-	-	0.012	0.010
2006q4	0.009	-0.012	-0.010	-0.016	-0.006	-0.014	-0.014	-0.012	-0.018	-0.019	-0.014	-0.015	-	-	…	-0.012
Mean	-0.012	0.000	-0.002	0.007	0.000	-0.001	-0.004	0.011	-0.001	-0.003	-0.005	-0.007	-0.008	-0.006	0.008	0.000

Notes:

… Missing data for some variables resulting in no scale economies estimates

-Not applicable due to mergers or late entry into industry

Source: Author's own computations, based on BoZ data

Appendix V: Iterated cost function estimation results

	Coefficient	Parameter	Standard error	t-statistic	p-value
Intercept	β_0	4.419	0.718	6.151	0.000***
$\ln(w_L)$	β_1	0.074	0.238	0.311	0.756
$\ln(w_K)$	β_3	1.069	0.241	4.439	0.000***
$\ln(tass)$	β_y	0.312	0.140	2.236	0.025**
$1/2\ (\ln(tass))^2$	β_{yy}	0.052	0.014	3.662	0.000***
$1/2\ (\ln(w_L)^2$	δ_1	0.063	0.035	1.783	0.075*
$1/2\ (\ln(w_K))^2$	δ_3	0.017	0.037	0.463	0.643
$\ln(w_L)\ln(w_K)$	δ_{13}	-0.057	0.032	1.779	0.075*
$\ln(w_L)\ln(tass)$	δ_{1y}	-0.011	0.016	-0.699	0.484
$\ln(w_K)\ln(tass)$	δ_{3y}	0.037	0.015	2.479	0.013***
t	θ_t	-0.005	0.011	-0.450	0.653
t^2	θ_{tt}	0.000	0.000	-0.309	0.758
$t\ln(tass)$	θ_{ty}	0.001	0.001	1.154	0.248
$t\ln(w_L)$	θ_{tL}	0.002	0.002	1.136	0.256
$t\ln(w_K)$	θ_{tK}	-0.004	0.002	-2.748	0.006***
Control Variables					
$\ln(BRANCH)$	ψ_1	0.042	0.012	3.449	0.001***
$\ln(RISK)$	ψ_2	0.012	0.008	1.606	0.108
$\ln(INTERMED)$	ψ_3	0.030	0.010	2.926	0.003***

Diagnostics

Equation	Obs.	Parameters	RMSE	R^2	p-value
Cost Function	388	17	0.146	0.995	0.000***
Labour input share	388	3	1083.940	0.698	0.000***
Capital input share	388	3	141.579	0.405	0.000***

Significance level: * p<0.10, ** p<0.05, *** p<0.01
Source: Author's own computations based on BoZ data