An Empirical Study of Bank Efficiency in South Africa Using the Standard and Alternative Approaches to Data Envelopment Analysis (DEA)

"Coert Erasmus, Daniel Makina
University of South Africa, Pretoria, South Africa
erasmcf@unisa.ac.za

Abstract: The paper investigates the efficiency of the major banks of South Africa using the standard and alternative approaches to Data Envelopment Analysis (DEA). The standard DEA approach measures efficiency utilising linear averages of outputs and inputs while the alternative DEA approach utilises nonlinear averages. Individual bank efficiency scores are estimated over the period 2006 to 2012, a period that allows analysis of the efficiency of the banks during the global financial crisis of 2008 to 2009. Under both approaches the majority of the major South African banks were observed to be DEA efficient, with the alternative approach improving the efficiency scores of those banks that were DEA inefficient under the standard approach. The global financial crisis did not affect the efficiency of the majority of the banks. Since the banks were DEA efficient prior the crisis, it could be argued that their efficiency was one of the contributory factors for their resilience during the global financial crisis.

Keywords: Bank efficiency; non-parametric method; DEA; intermediation approach; South Africa

1. Introduction

Despite there being extensive research in measuring efficiency in financial institutions, there is no generally accepted best method (Frimpong, 2010). Early research utilised conventional financial ratios such as return on equity (ROE) and return on assets (ROA). Later studies utilised parametric and non-parametric approaches. Parametric methods include the stochastic frontier and thick frontier approaches while the non-parametric methods comprise the Free Disposal Hull (FDH) approach, the Index Numbers (IN) approach and the Data Envelopment Analysis (DEA) approach (Tahir & Yusof, 2011; Berger & Humphrey, 1997). As to what approach best measures bank efficiency is still an unresolved question though the DEA approach pioneered by Farrell (1957) is increasingly being preferred (Ferrier & Lovell, 1990; Berger & Humphrey, 1997). Most studies on bank efficiency have focused on developed economies especially the USA and Europe and very little is known about developing countries. With regard Sub-Saharan African banks, Chen (2009) attributes the scant knowledge to the region having fewer banks, low level of financial development, limited market activity and lack of quality data. Nevertheless, some middle-income countries like South Africa, the subject of this study, have relatively complex financial systems, with well-developed banks and available data that allow insights into their operation and efficiency to be studied. Illustrating the significant role banks play in the economy, Botha and Makina (2011) report that the gross value added of the financial sector in South Africa is about 10.5% of GDP in which bank assets alone constitute 127% of GDP.

A study of bank efficiency is important because it is generally acknowledged that financial sector development is a crucial ingredient for economic growth. Honohan and Beck (2007) emphasize that the services provided by the financial sector of mobilization of savings and facilitating transaction services and risk management services are critical for development. Therefore, the operational efficiency of banks is crucial for the smooth development of an economy. In fact, the long-term viability of banks is linked to their levels of efficiency. The study of bank efficiency is also helpful in locating sources of inefficiencies to enable stakeholders to initiate reforms and design suitable strategic measures especially in developing economies (Chen, 2009). The choice of South Africa as a unit of study is driven from the stylised fact that the country has a well-developed banking sector similar to those in the developed countries yet its efficiency has not been extensively researched. To our knowledge all studies on bank efficiency carried in South Africa have used either the stochastic frontier approach or the standard DEA approach on data, typically from 1999 to 2009 (Mlambo & Ncube, 2011; Van Heerden & Van der Westhuizen, 2008; Okeahalam, 2006; O’Donnell & van der Westhuizen, 2002). Our approach differs in several respects from these studies. First, previous studies assessed the efficiencies of banks over periods before the 2007-2009 global financial crises whereas our approach assesses efficiencies over a horizon that includes both the
pre- and post-crisis periods. Second, previous studies focused on a sample of banks that included small and big banks whereas our approach focuses on the five big banks that could be considered by authorities as too big to fail. In any case the sample of the five big banks is representative of the South African banking sector as they represent over 85% of the market. Hence conclusions arrived at regarding efficiency would be representative of the South African banking sector. Third, in addition to using the standard approach, we also estimate technical efficiency of the banks using the alternative DEA approach that utilises log-linear data. The rest of the paper is structured as follows. Section 2 discusses relevant literature and studies that have been carried out in South Africa. Section 3 describes the research methodology chosen including its pros and cons. Section 4 presents and discusses the empirical results. Finally, section 5 concludes.

2. Literature review

Significant changes in the banking sector such as deregulation, globalization, financial innovation and technological progress have reduced costs of information processing and thus enabling bank management to improve resource allocation. Academic research has accordingly focused on measuring bank efficiency. According to Usman et al. (2010) bank efficiency is the best level of output reached without any changes to the amount of input. Efficiency is generally categorised in to two components – allocative and technical efficiency (Xu, 2011; Neube, 2009). However, these components have since been expanded to include economic and scale efficiency. According to Farrell (1957), technical efficiency refers to a decision making unit’s ability to produce the maximum outputs at a given level of inputs (known as the output orientation) or ability to use the minimum level of inputs at a given level of outputs (known as input orientation). Thus it can be measured within two main frameworks: input-oriented and out-oriented frameworks. Technical efficiency in a banking environment measures the ratio between the anticipated level of outputs and the actual level of inputs (Xu, 2011). Effective management of assets and profit reinvestment would return higher technical efficiency for banks, while excessive operating expenditure reduces the technical efficiency (Assaf et al., 2011; Van Heerden & Van der Westhuizen, 2008). A bank is technically efficient when it operates on the production efficiency frontier. The focus of this paper is to measure technical efficiency.

Allocative efficiency measures the ability of optimising the outputs with the technology available to the institution or minimising inputs by decreasing current costs (Hon, Tuck & Yu, 2011; Koutsomanoli-Filippaki et al., 2011; Xu, 2011; Brissimis et al., 2010; O’Donnel & Van der Westhuizen, 2002). Economic efficiency or cost efficiency is the product of the allocative and technical efficiency (Hon et al., 2011; Brissimis et al., 2010). Banks which is cost efficient produces a level of output at the lowest possible cost with the price and technology constraint at hand. They are viewed as scale efficient when a single input produces several outputs. An investigation by Humphrey (1990) and Berger and Humphrey (1991) showed that banks with a lower level of assets often returned increasing returns to scale compared to banks with a higher level of assets which were observed to exhibit decreasing scale returns. A number of researchers focus on profit efficiency (Assaf et al., 2011; Fethi & Pasioras, 2010). Assaf et al. (2011), for instance, observe that banks with higher levels of efficiency yield lower net profit margins compared to their less efficient counterparts. Several approaches can be used to determine each type of efficiency. Broadly, they are classified into two categories, namely, parametric and non-parametric methods to measure the levels of efficiency (Berger and Humphrey, 1997); Frimpong, 2010). Researchers still disagree on the most suitable method to measure efficiency as each method holds certain advantages and disadvantages. In the literature the most favoured efficiency measurement techniques for parametric and non-parametric analysis are respectively the Stochastic Frontier Analysis (SFA) and Data Envelope Approach (DEA). Usman et al. (2010) concluded that there is no superior method exists to measure bank efficiency. However, the DEA has been found to have a number of advantages over other methods (Nigmonov, 2010; Khankhoje and Sathye, 2008). First, it does not require specification of any functional relationship between inputs and outputs or a priori specification of weights of inputs and outputs. Second, it easily accommodates multiple inputs and outputs for which are the norm for the banking sector. Third, it is suitable for measuring the efficiency of firms that lack competitive prices as could be the case of a concentrated banking sector like that of South Africa.

The definition and measurement of inputs and outputs in the banking environment remain a contentious issue (Xu, 2011; Brissimis et al., 2010; Usman et al., 2010; Nigmonov, 2010). Literature distinguishes two fundamental approaches – the production and intermediation approaches – for which there no consensus among researchers. According to Sathye (2003) Khankhoje and Sathye (2008), most studies in the DEA
literature utilises the intermediation approach and within this approach the exact set of inputs and outputs used depend on data availability. Ferrier and Lovell (1990) observe that measuring efficiency using the intermediation approach is concerned with evaluating the viability and the cost of banking. Several authors integrated various variables into the measurement as inputs and outputs (Hon et al., 2011; Koutsomanoli-Filippaki et al., 2011; Assaf et al., 2011; Sealey & Lindley, 1977). With an intermediation approach deposits is the most important input, other inputs include interest expenses and non-interest expenses. Output variables include loans to customers, interbank loans, interest income, non-interest income, staff salaries, number of employees and fixed assets. However, researchers have debated whether deposits should be considered as an input or output (Berger & Humphrey, 1997). In recent literature, Holod and Lewis (2011) concluded that there is still no correct set of input and output variables defined for the intermediation approach. Nevertheless, these researchers accept that banks are intermediaries between the surplus economic units and the deficit economic units (Holod & Lewis, 2011; Berger & Humphrey, 1997). Advocates of the intermediation approach (e.g. Brissimis et al., 2010; Van Heerden & Van der Westhuizen, 2008) reason that the intermediation approach is superior to the production approach and defend their position with three reasons. Firstly, the approach includes the interest expenses, which expenses could account for more than half of the total cost of a bank. Secondly, the total costs are minimized, and not only the production costs. Thirdly, the data quality is unparalleled.

The production approach defines the traditional inputs such as land, labour and capital as variables used to produce the outputs (Usman et al., 2010; Ncube, 2009). Sathye (2003) observes that the traditional inputs generate the output of deposits and that this is the single most important difference between the two approaches. The production approach also has a drawback as it does not account for the mitigation of financial risk that a bank incurs when engaging in loan transactions (D’Souza & Lai, 2004). A number of studies indicate interdependency between bank efficiency and economic development (Koutsomanoli-Filippaki et al., 2011; Staub et al., 2010; Usman et al., 2010). From these studies it is observed that an efficient financial sector is important for both developed and developing countries. European countries acknowledged that an efficient banking sector is essential for economic recovery after the 2008 global financial crisis. Research also suggested that developing countries with a more efficient financial sector enjoyed increased economic growth. While Sub-Saharan African countries generally experience similar challenges in the banking environment as other countries, research on bank efficiency is sparse (Chen, 2009). South Africa is described as an ideal country for bank research as it has a banking sector similar to those found in developed economies (Akinboade & Makina, 2009). According to Ncube (2009), the challenges South African banks face includes foreign and domestic competition, rising operational costs and regulatory changes. Some authors argue that political interference in the banking sectors of African countries could negatively impact the banking efficiency (Barros et al., 2012). However, not much research has been undertaken to investigate these issues. The few studies on bank efficiency include those of Okeahalam (2004) focusing on Uganda and Botswana; Okeahalam (2006) focusing on South Africa and using the Bayesian stochastic frontier analysis; and studies by Mlambo & Ncube, (2011), Van Heerden & Van der Westhuizen (2008) and O’Donnell & van der Westhuizen (2002) focusing on South Africa and utilising the DEA approach. None of these studies considered the effect the global financial crisis could have had on the efficiency of the banking sector in South Africa. This paper attempts to explore efficiency scores pre- and post-crisis.

3. Methodology

Approach and Data: We utilise the standard DEA, a non-parametric approach that does not require the specification of the functional form of the production function, to assess the technical efficiency of five big banks in South Africa which are treated as distinct decision making units (DMUs). DEA is a linear programming technique initially developed by Farrell (1957) to measure the different DMUs based on single input-output variables. This technique was extended by Charnes, Cooper and Rhodes (1978) to evaluate the efficiency of public sector non-profit organisations using multiple inputs and outputs under the assumption of constant returns to scale. It was further extended by Banker, Charnes and Cooper (1984) to include efficiency under the assumption of variable returns to scale. Sherman and Gold (1985) were the first to apply the technique to banking. The main objective of DEA is to determine which banks are operating on their efficient frontier and those which are not. If the bank’s input-output combination lies on the DEA frontier, the bank is viewed as efficient, and if the bank’s input-output combination lies inside the frontier, it is considered inefficient. The standard DEA approach is premised on a linear programming formulation (Ragsdale, 2007). In line with the approach adopted in the literature, we utilise the intermediation approach in defining the inputs and outputs of our sample of banks (Khankhoje &
The intermediation approach considers financial institutions as primarily intermediaries channelling funds between surplus units (savers) and deficit units (borrowers). After having regard to availability of data, the inputs and outputs for each bank (all publicly quoted on the Johannesburg Securities Exchange (JSE)) - Standard Bank, Barclays (ABSA) Bank, FirstRand Bank, Nedbank and Capitec Bank - we utilise inputs and outputs are listed in Table 1 below. The secondary data of the five banks' inputs and outputs covering the period 2006 to 2012 were obtained from Bloomberg and McGregor BFA databases.

### Table 1: Inputs and Outputs under the Intermediation Approach

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposits</td>
<td>Loans and Overdrafts</td>
</tr>
<tr>
<td>Other liabilities</td>
<td>Non-interest income</td>
</tr>
<tr>
<td>Shareholders' equity</td>
<td></td>
</tr>
<tr>
<td>Staff costs</td>
<td></td>
</tr>
<tr>
<td>Non-interest expense</td>
<td></td>
</tr>
<tr>
<td>Fixed assets</td>
<td></td>
</tr>
</tbody>
</table>

Notwithstanding that the standard DEA is a powerful tool, it suffers from the manner in which outputs and inputs are aggregated. According to Shirvani et al. (2011), the standard DEA model assumes that all outputs and inputs are perfect substitutes for each other when this is not the case. For instance, bank employees and branches are treated as perfectly substitutable so that a bank can simply add new employees to existing branches instead of increasing branches. Furthermore, its use of linear averages makes relevant efficient ratios nonlinear fractions so that direct application of linear programming is not possible. Hence, the model is forced to standardise the denominator of the efficiency ratio (the weighted average input) to equal one. Shirvani et al. (2011) offer an alternative DEA approach in which the use of nonlinear (geometric) weighted averages produces a log-linear relationship among the relevant variables and thus making it possible to directly utilise linear programming for optimisation purposes.

**Model Specification:** Under the standard DEA approach, the efficiency of bank $i$ is defined as:

$$\text{Efficiency of bank } i = \frac{\text{weighted sum of bank } i\text{'s outputs}}{\text{weighted sum of bank } i\text{'s inputs}}.$$ 

Supposing we have $N$ DMUs (five banks in our case), each with $n$ inputs and $m$ outputs, the DEA relative efficiency score of a given bank is obtained by solving the following linear programming model:

Maximize $e_s = \frac{\sum_{i=1}^{m} u_i y_{is}}{\sum_{j=1}^{n} v_j x_{js}}$ \hspace{1cm} \text{For } i = 1, 2, \ldots, m; j = 1, 2, \ldots, n \hspace{1cm} [1]$

Subject to:

$$\frac{\sum_{i=1}^{m} u_i y_{ir}}{\sum_{j=1}^{n} v_j x_{jr}} \leq 1 \hspace{1cm} \text{For } r = 1, 2, \ldots, n$$

$u_i > 0; v_j > 0$

Where:

- $e_s$ = the efficiency score;
- $y_{is}$ = the amount of the $i$th output produced by the $s$th bank;
- $x_{js}$ = the amount of the $j$th input used by the $s$th bank;
- $u_i$ = the output weight;
- $v_j$ = the input weight;
- $x_{jr}$ = the amount of input $j$ utilised by the $r$th DMU; and
- $y_{ir}$ = the amount of output $i$ produced by the $r$th DMU

Since the objective function in equation [1] is in fractional form, it needs to be reformulated to a linear equation [2] as follows in order to run DEA on a standard linear program package:

Maximize $e_s = \sum_{j=1}^{n} v_j x_{js}$ \hspace{1cm} [2]

Subject to:

$$\sum_{i=1}^{m} u_i y_{is} = 1$$

$$\sum_{i=1}^{m} u_i y_{is} - \sum_{j=1}^{n} v_j x_{jr} = 0$$

$u_i, v_j > 0$
The DEA establishes a benchmark efficiency score of unity that no DMU (bank) can exceed. Thus, an efficient score of one satisfies the necessary condition of being DEA efficient, and that below one indicates being DEA inefficient.

On the other hand, the maximization under the alternative DEA is specified as:

Maximize:
\[
\log e_s = \sum_{j=1}^{n} u_j \log y_{ir} - \sum_{j=1}^{m} v_j \log x_{jr}
\]  

Subject to:
\[
\log e_s \leq 0 \text{ for } j = 1, 2, \ldots, n
\]
\[
\sum_{j=1}^{n} u_j + \sum_{j=1}^{m} v = 1
\]

We proceed to estimate the standard DEA and the alternative DEA using inputs and outputs of the five banks covering the period 2006-2012 and compare the results of the two approaches. Empirically, the alternative DEA approach has been shown to exhibit less inefficiency among banks than the standard DEA approach (Tahir et al., 2013). It would be interesting to see whether the results of this study which is using longer dated data would be consistent with those of Tahir et al. (2013) who only used a three year data set.

4. Findings and discussion

Table 2 below presents the standard DEA efficiency scores of individual banks over the period 2006 to 2012.

<table>
<thead>
<tr>
<th>DMU</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barclays Africa (ABSA)</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Capitec Bank</td>
<td>72.9%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Firstrand Bank</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>83.4%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Nedbank</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Standard Bank</td>
<td>99.1%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>87.9%</td>
<td>93.1%</td>
<td>85.7%</td>
</tr>
</tbody>
</table>

During the period 2006 to 2012 two banks – Barclays Bank and Nedbank – exhibited DEA efficiency throughout. Analysed year by year, in 2006 only two banks – Capitec Bank and Standard Bank were DEA inefficient and the rest were DEA efficient. In 2007 to 2008 all the banks were DEA efficient and in 2009 at the height of the global financial crisis only one bank – Firststrand Bank- was DEA inefficient. From 2010 to 2012 four banks –Barclays Bank, Capitec Bank, Firststrand Bank and Nedbank –were DEA efficient whereas Standard Bank was DEA inefficient throughout this period. Table 3 below shows efficiency scores under the alternative DEA approach.

<table>
<thead>
<tr>
<th>DMU</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barclays Africa (ABSA)</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Capitec Bank</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Firststrand Bank</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>98.5%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Nedbank</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Standard Bank</td>
<td>99.8%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>99.9%</td>
<td>99.2%</td>
<td>99.1%</td>
<td>98.7%</td>
</tr>
</tbody>
</table>
The alternative DEA approach shows increased overall efficiency scores as compared to the standard DEA approach. During the period 2006 to 2012 three banks – Barclays Bank, Capitec Bank and Nedbank – exhibited DEA efficiency throughout as opposed to only two banks according to the standard DEA approach. Analysed year by year, in 2006 only one bank as opposed two with standard approach – Standard Bank – was DEA inefficient and the rest were DEA efficient. Though DEA inefficient, the efficiency score of Standard Bank improved under the alternative DEA approach. In 2007 to 2008 all the banks were DEA efficient and in 2009 at the height of the global financial crisis only two banks – FirstRand Bank and Standard Bank – was DEA inefficient. However, efficiency scores improved under the alternative approach. From 2010 to 2012 four banks – Barclays Bank, Capitec Bank, FirstRand Bank and Nedbank – were DEA efficient whereas Standard Bank was DEA inefficient throughout this period but with improved efficiency scores under the alternative approach. These results are consistent with the observations of Tahir et al. (2013) and Shirvani et al. (2011). It is noteworthy that the global financial crisis did not have a significant impact on the technical efficiency of the major banks in South Africa. While technical efficiency could have been one of the contributing factors, a number of reasons have been advanced for the resilience of South African banks (National Treasury Policy Document, 2011).

First, South Africa is endowed with a sound framework for financial regulation. Regulators have generally shied away from a "light-touch" approach, that is, the belief that the financial sector can effectively regulate itself. For instance, excessive credit extension is curbed through effective legislation in the form of the National Credit Act that ensures that banks lend responsibly to households so that they are not overly indebted. The Registrar of Banks also limited credit extension by banks through raising capital adequacy requirements and setting conservative leverage ratios. Second, South African banks have conservative practices so that there isn’t much securitisation and derivatives trading relative to developed markets that take place. Furthermore, the South African experience of the small banking crisis in 2002 and the adoption and implementation of Basel II Capital Accord in 2008 ensured improved risk management practices. Third, South Africa has prudential regulation of foreign exposure that limited overall foreign risk. That limits are placed on the extent of exposure to foreign assets by institutional investors and banks ensured no exposure to sub-prime investments. Fourth, South Africa requires registered banks to be subsidiaries of the domestic or foreign parent company so as to ring-fence their assets and liabilities in the event that the parent company is in distress. Furthermore, for those banks that choose to list on the JSE, the listing requirements ensure transparency, rigorous disclosure standards and high standards of corporate governance.

Finally when comparing the results to previous studies of bank efficiency in South Africa that made use of the DEA approach we observe some unique attributes of this paper. This paper presents the individual efficiency scores of the five major banks listed on the JSE – Barclays Africa, Capitec bank, FirstRand bank, Nedbank and Standard Bank – that represents over 85% of the South African banking sector. The earlier research by O’Donnell and Van der Westhuizen (2002) determined the efficiency of a single bank at branch level. While, Van Heerden and Van der Westhuizen (2008) investigated the efficiency of only a single unnamed bank from one of the major banks. The authors did a monthly analysis of the efficiency scores. The results from the unnamed bank also yielded high levels of efficiency, supporting the results obtained in this study as some results were similar. Lastly Mlambo and Ncube (2011) measured the efficiency of an unbalanced panel of 26 banks and not individual banks as this paper. The data used was up to the year 2009 and a rising trend of efficiency was observed. All prior studies did not attempt to link efficiency to the global crisis as this paper has done.

5. Conclusion

The paper has investigated the efficiency of the major banks of South Africa using the standard and alternative approaches to Data Envelopment Analysis (DEA). The standard DEA approach measures efficiency utilising linear averages of outputs and inputs while the alternative DEA approach utilises nonlinear averages. Individual bank efficiency scores are estimated over the period 2006 to 2012, a period that allows analysis of efficiency of the banks during the global financial crisis of 2008 to 2009. Under both approaches the majority of the major South African banks were observed to be DEA efficient, with the alternative approach improving the efficiency scores of those banks that were DEA inefficient under the standard approach. The global financial crisis did not affect the efficiency of the majority of the banks. Since the banks were DEA efficient prior the crisis, it could be argued that their efficiency was one of the contributory factors for the resilience the South African banks during the global financial crisis.
References


