DOI: 10.1111/saje.12364

SPECIAL ISSUE ARTICLE

Measuring the unit cost of financial intermediation in South Africa: A measure of bank productivity

Keaoleboga Mncube | **Nicola Viegi**

University of Pretoria, Pretoria, South Africa

Correspondence
Keaoleboga Mncube, University of Pretoria, Pretoria, South Africa.
Email: keaolebogamncube@gmail.com

Abstract

This paper presents evidence on productivity growth in the South African banking industry in the last 30 years. The productivity measures we construct shed light on whether the development, increased contribution and influence of the banking industry have translated into lower cost of intermediation and improved productivity and efficiency of the banking sector. Our results show that there is no apparent trend in the unit cost of intermediation in the period 1993 to 2019, which is indicative of constant productivity.

KEYWORDS

bank output, banking sector, income share, productivity

JEL CLASSIFICATION

E24, G21, N27

1 | INTRODUCTION

The South African banking sector plays a primary role in financial intermediation[1] and is well developed, large and deep. For instance, credit intermediated by the sector to the private non-financial sector as a percentage of Gross Domestic Product (GDP) is consistently around 60%, in line with many advanced economies.[2] Assets of the sector as a percentage of GDP peaked at over 80% before the 2008 global financial crisis (GFC), relative to the United States (US) with bank assets as a percentage of GDP peaking at 60% before the GFC.[3] The sector has also grown twice as rapidly as the services sector and more than three times as fast as the manufacturing sector (Rashid, 2011). These facts point to the importance of the banking sector in the South African economy.

[1]Bazot (2017) defines financial intermediation as banks' activity of pooling risks, reducing asymmetries that impede transfer of funds and provision of liquidity.
[2]Figures obtained from Bank for International Settlements (BIS) available at https://stats.bis.org/statx/srs/table/f2.4.
[3]Figure is obtained from World Bank, Deposit Money Bank Assets to GDP for South Africa [DDDI02ZAA156NWDB], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DDDI02ZAA156NWDB.

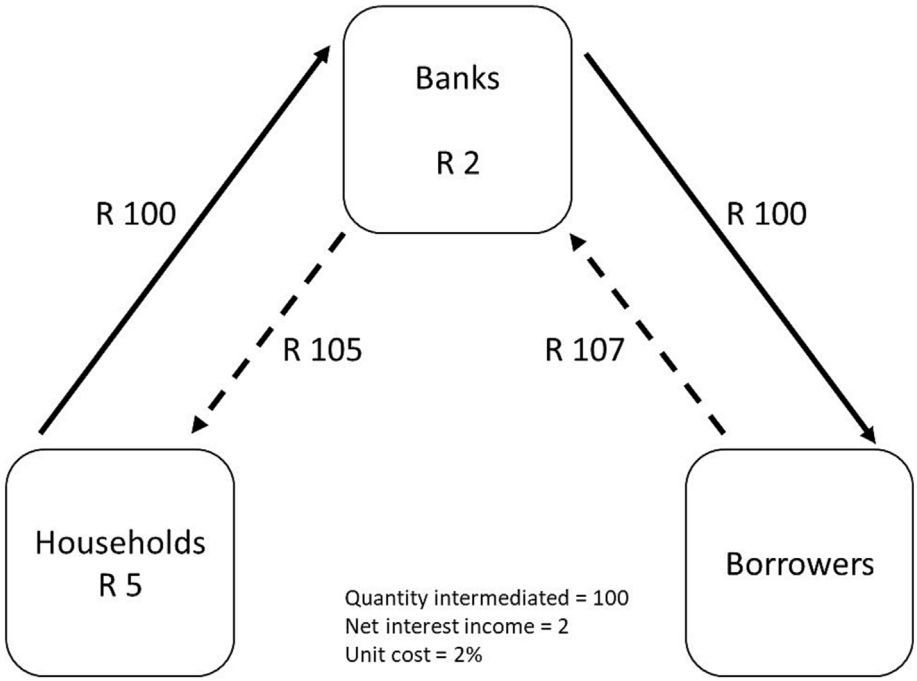

FIGURE 1 Simplified model of financial intermediation. *Note*: This figure shows a simplified model of financial intermediation. *Source*: Philippon (2015).

At the same time, the increasing dimension and importance of the banking system is not in itself an indicator of efficiency of financial intermediation. The cost of financial intermediation is another important channel through which financial development enhances productivity and growth in the economy.[4]

This paper therefore reviews evidence on productivity growth in the banking industry over the past 30 years and investigates whether the development, increased contribution and influence of the sector have translated into lower cost of intermediation, implying improved productivity and efficiency.

In doing so, we aim to construct measures of income for the banking system, quantity of intermediated assets and of unit cost of financial intermediation. In this, we follow the work of Philippon and Reshef (2013), Philippon (2015) and Bazot (2017) that present the same evidence for the US and several European countries.

To interpret our measures of bank productivity, we use a simple model of financial intermediation illustrated in Figure 1, adapted from Philippon (2015). In this model, banks provide lending and deposit services. To illustrate, we assume that depositors want to deposit R100 with a 5% deposit rate and obtain R105 in return, while borrowers want to borrow R100 at a 7% loan rate, thereby repaying R107. Banks reduce information asymmetries and play an intermediary role between depositors and borrowers. The R100 is viewed as the quantity of the intermediated asset, while the R2 difference accrued to the bank is the interest income that compensates for the cost of intermediation such as the cost of monitoring and managing the risk of the loan. In essence, the unit cost of intermediation is R2 or 2%.

Therefore, as the banking industry becomes more productive, it should cost less to maintain and create an intermediated asset.

[4]The role that the cost of financial intermediation plays in the efficient allocation of resources, and in favouring economic growth is well established in the literature, see, for example, Levine (1997) for a classic reference and Greenwood et al. (2010) and Buera et al. (2011) for more recent ones. The cost of financial intermediation plays an important role also in macroeconomic models of the business cycle as in Curdia and Woodford (2016).

Evidence from Philippon and Reshef (2013), Philippon (2015) and Bazot (2017) shows that the unit cost of intermediation has not declined since the 1950s in the

US and European banking industries. In other words, despite the development and innovation in the finance industry, financial services are still expensive, and in fact, the finance industry of 1900 was just as able as the finance industry of 2010 to produce loans, bonds and deposits, even cheaply.

Our results show three main stylized facts:

- Output, gross value added (GVA) and remuneration in the banking sector have grown significantly until the GFC. After 2007, remuneration and GVA have been stagnant against an increase in bank output, indicating a decline in the value of bank output.
- There is no apparent trend in our measure of bank productivity, the unit cost of intermediation.[5] The unit cost rises post-1999 until 2002, which is a period characterised by various mergers and consolidation in the banking industry, indicating a potential reduction in competition.[6] Post-2002, the cost ratio of core banking services (credit and deposits) declines back to the 1993 levels, as the overall cost of finance.
- Labour productivity increases significantly before the GFC but has since halved in the 10 years following the GFC when the banking industry saw a large expansion in employee numbers, not matched by a contemporaneous increase in bank output. This is partly related to the increased investment in regulatory and compliance activities to satisfy the new regulatory environment developing after the GFC.

Therefore, despite major improvements in the South African banking industry with regards to technological advances, the cost of creating and maintaining intermediated assets has been constant. A further implication of the mixed trend in the cost ratio is that the improvements have not been passed to the end users of banking services (i.e. households and firms) in the form of lower cost of intermediation.

Studies on the productivity and efficiency of the South African banking sector have focused on correlating productivity and efficiency with competition in the sector or with concentration and market power. For instance, Okeahalam (2006) analyses the production efficiency of bank branches, in the context of a concentrated banking system such as in South Africa. Verhoef (2009) correlates bank concentration to the efficiency ratio of banks and finds that banks' efficiency declined post-2002, with the period correlating with increased concentration in the sector. Maredza and Ikhide (2013) measure the impact of the GFC on the efficiency and productivity of the banking sector, with their results indicating that total factor productivity (TFP) efficiency was 16.96% lower during the crisis period compared to the pre-crisis period. Mlambo and Ncube (2011) investigate the evolution of competition and efficiency in the sector, and their results show that the number of efficient banks was falling between the 1999 and 2008 period. Notably, the approaches in the above works in determining and defining efficiency and productivity are based on econometric approaches, apart from Verhoef (2009).

The rest of the paper is organised as follow. Section 2 discusses the data and measurement of income share and output of the banking industry. Section 3 discusses the results and an additional measure of labour productivity. Section 4 concludes and suggest further research analysing the impact of prudential regulation in reducing the dynamism and efficiency of the sector.

2 | DATA AND METHODOLOGY

2.1 | Measuring financial income and output

In this section, we describe the data used to construct the bank productivity measures. This is based on three correlated but different indicators of the size of the banking sector. The source of the data is in Table A1.

[5]The terms unit cost of intermediation and cost ratio of finance are used interchangeably in the paper.

[6]The banks that had liquidity problems and thereby placed under curatorship and those that were eventually liquidated include African Bank, Islamic Bank, Regal Treasury Bank and Saambou Bank, among others. The biggest merger occurred in 1997 between First National Bank, Rand Merchant Bank, Southern Life assurance company and Momentum Life assurance company to form the First Rand Group.

The first measure is the income share of the banking industry, which gives an idea of the economic weight of the sector over time. Following Philippon (2015), we use the ratio of GVA to GDP and the ratio of labour compensation in banking to aggregate compensation in the economy as measures of income share. As outlined by Haldane et al. (2010), GVA is the sum of profits and wages in the banking sector, which we divide by GDP to obtain a measure of income share. Similarly, labour compensation of employees captures wages, salaries and bonuses, which we divide by aggregate labour compensation in the economy.

We also describe the size of the banking sector in terms of output produced. According to Bazot (2017), bank output accounts for all services provided by banks, with an intermediated asset defined as an asset that provides a financial service to non-financial industry customers. Therefore, we define the proxy for bank output or quantity of intermediated assets[7] as the sum of credit to the private non-financial sector, liquidity services, equity intermediation and public sector marketable debt. Credit to the private non-financial sector includes overdrafts, loans and advances extended by banks to households and non-financial sector firms. These are measured on the asset side of banks' balance sheets. We use deposits to proxy for liquidity services, measured on the liabilities side of banks.

Stock market capitalization is used as a proxy for equity intermediation, while public sector marketable debt includes financial debt instruments of government institutions at all levels of general government and public sector corporations. It is worth noting that the development and evolution of banking industries have seen banks taking on other non-traditional banking services such as asset and wealth management. However, due to difficulties in measuring such services, we limit and define bank output as described above.

Having described income share and output of the banking sector, the measure of bank productivity is computed as income share divided by bank output.

3 | RESULTS

3.1 | Income share and Bank output

Figure 2 shows the evolution of the measures from 1993 to 2019. Income share notably declined between 1997 and 1999, which is a period of significant structural change in the banking sector as described earlier. Post-2002, the industry's contribution expands robustly reaching a peak of over 4% before declining and flattening post the GFC. This dynamic is highly correlated with the overall growth trajectory of the economy.

Figure 2 also shows that the relative remuneration measure follows the GVA path, with continuous growing importance until the GFC and stagnation afterwards. Philippon and Reshef (2013), Philippon (2015) and Bazot (2017) find similar in- creasing trends in income share of finance in the United Kingdom, Japan, Canada and the US, respectively.

Although it is not the objective of the paper to outline reasons behind the upward trend of income share, it is worth noting an important aspect. The rise of income share post-2002, particularly the 2002 to 2006/2007 period, coincides with robust economic growth, coupled with strong wage, credit and deposit growth. This suggests that rising income share of finance is linked to South Africa's economic growth.[8]

The evolution of bank output is also shown in Figure 2, with evident expansion post-2002, coinciding with the 2002 to 2006/2007 economic growth period. After the crisis and a brief contraction of bank output, the sector has started growing again in proportion to the economy. It is noticeable how in the last 10 years, the growth of bank total output is not matched by a similar growth in income share.

[7]The two terms are used interchangeably in the paper.
[8]GDP growth peaked at 5% in 2005, the highest to date, while wages in banking, credit and deposit growth all peaked at over 10%.

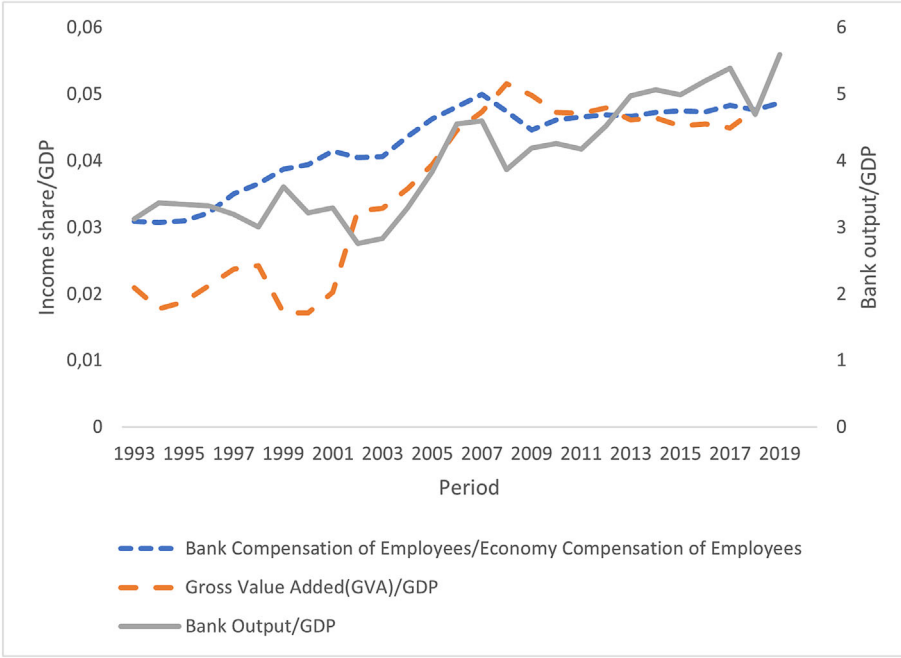

FIGURE 2 Financial income and output. *Note*: This figure shows measures of income share of the South African banking industry (finance), which are gross value added by the banking industry, as a percentage of GDP and relative remuneration, which is the ratio of compensation of employees in banks to overall compensation of employees in the economy. It also shows quantity of intermediated assets or bank output. *Source*: Quantec and authors' calculations.

3.2 | Unit cost of intermediation

We proceed to calculate the measure of productivity, computed as income share divided by bank output. It measures the cost of creating and maintaining an intermediated asset, given a set of inputs (Philippon, 2015). Further and in line with Philippon and Reshef (2013) and Philippon (2015), we compare the unit cost of intermediation for different measures of bank output where we use credit to the non-financial private sector and deposits as bank output and an attempt to encompass a broader measure of bank output, as defined in Section 2.

Figure 3 shows the evolution of the unit cost of financial intermediation. Notably, the cost ratio falls leading up to 1999 and thereafter rises to over 1% as a percentage of GDP. The rising cost ratio post-1999 is driven by rising income share relative to bank output, which is indicative of the notion that how much society is paying for banking services is not matched by what society is obtaining from banks. In addition, Philippon (2015) argues that a possible explanation for the rise in the cost of intermediation in the US is increased concentration in the banking sector from 1998. Therefore, a similar argument can be made for South Africa as the rise in the cost of intermediation occurs during a period of increased concentration in the banking sector. The finding is also complimentary to the work of Verhoef (2009) who shows that efficiency of banks in South Africa declined post-2001 until 2006.

Post-2002, the cost ratio declines until 2006, rises briefly until 2008 and eventually flattens post-2008 to almost similar levels in 1993. The declining cost ratio post-2002 signals improved productivity as the increase in bank output outpaces what society pays for such bank services.

However, a puzzling implication of the flat cost ratio post-2008 is its similarity to levels almost experienced in 1993. This implies that improvements in the banking industry such as technological improvements have not been passed to the end users of banking services. The expectation is that such improvements and innovations would improve productivity and imply lower cost of banking services for

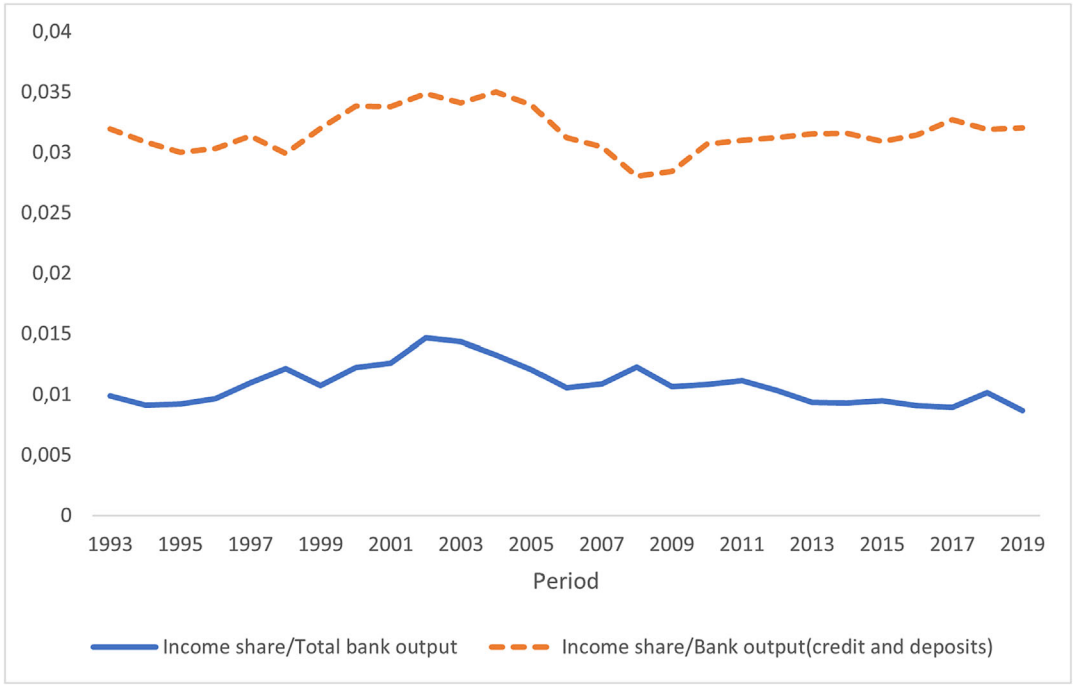

FIGURE 3 Cost ratio of finance. *Note*: This figure shows the evolution of the cost ratio of finance in South Africa, which is a measure of productivity in the banking sector. It is computed as income share, divided by bank output. *Source*: Authors' calculations, Quantec and Stats SA Quarterly Employment Statistics (QES).

customers. Despite using different measures of bank output and a longer period of analysis, Philippon and Reshef (2013) and Bazot (2017) find similar mixed trends in the cost ratio for the United Kingdom, France and Germany. A similar finding is observed for the US when bank output is comprised of loans only.

We further relate the unit cost of intermediation to an existing measure of labour productivity, which is an index of labour productivity for banks. It is measured as output per unit of labour.[9] The measure is also used by Maredza and Ikhide (2013) in their analysis of the impact of the 2008 financial crisis on efficiency and productivity in the banking system. The intuition behind the correlation is as follows. Output per unit of labour analyses the use of labour input to produce output, while the unit cost of intermediation considers income accruing to production factors. Therefore, as the banking industry becomes more productive, reflected by an increase in the labour productivity index, it should cost less to maintain and create an intermediated asset, reflected through a decline of the unit cost of intermediation.

Figure 4 shows an improvement in labour productivity from 2002 until 2007. In the same period, the cost ratio falls as described above and in Figure 3. Post-2008, labour productivity falls due to the expansion of labour input, as shown in Figure 4 by the rise in the number of employees in banks, relative to bank output. Our measure of productivity, the unit cost of intermediation, is constant due to flat income share post-2008.

A further puzzling implication from our analysis is the rise in the number of employees as shown in Figure 4, which is not matched by a rise in bank output.[10] Therefore, could the expansion in labour and

[9]The definition is in line with the International Labour Organisations' definition of labour productivity.

[10]The rise in bank output post-2008 shown in Figure 2 is attributed to the rise in the intermediation of equity. Credit extension and deposit services fall and flatten in the same period to levels experienced before the 2003/2004 credit boom.

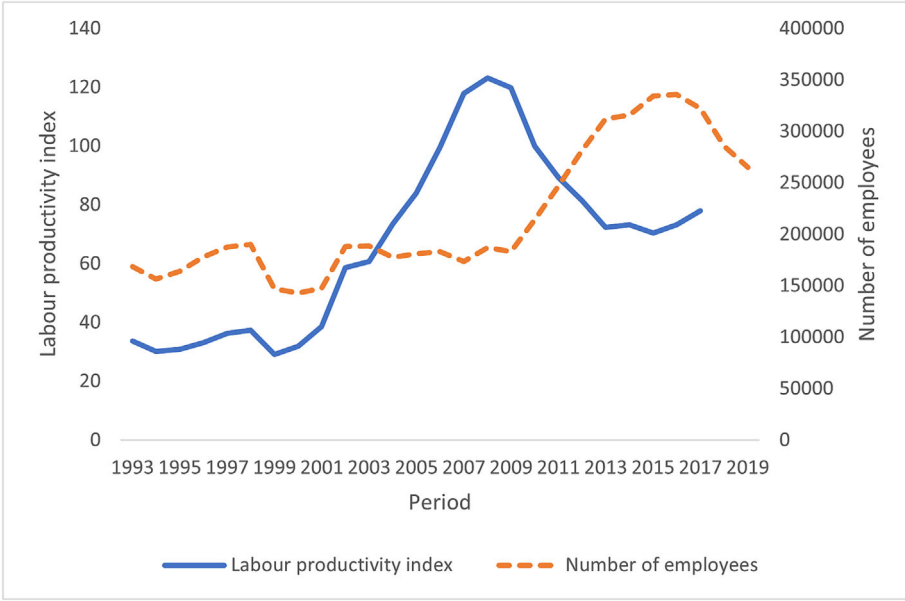

FIGURE 4 Labour productivity index and employees in banks. *Note*: This figure shows the labour productivity index in the South African banking industry, defined as output per unit of labour. It also shows the number of employees in the banking industry. *Source*: Authors' calculations, Quantec and Stats SA Quarterly Employment Statistics (QES).

subsequent decline in labour productivity, coupled with constant bank productivity as measured by the unit cost of intermediation, be explained by changes in bank regulation?

The post-2008 period is significant as further regulatory changes in the banking industry were introduced following the GFC. However, banks in South Africa and reports by other financial institutions raised concerns relating to the intrusive and costly nature of regulation. For instance, the PricewaterhouseCoopers (PWC) Major Banks' analysis report[11] and the Financial Services Sector Assessment Report[12] indicates that the introduction of regulation has seen banks in South Africa subsequently investing in regulation and compliance staff as well as system enhancements needed to implement and meet regulatory report requirements. Consequently, headcount staff and costs continue to increase, accounting for over 55 percent of major banks' total operating expenses.

Chief Executive Officers (CEO's) of major banks in South Africa have in addition, raised concerns relating to efforts spent on regulatory compliance, which requires significant time and focus to adapt to and implement.[13] Furthermore, the Financial Services Sector Assessment Report indicates that changes in bank regulation are occurring at a rapid rate, with concerns that they might be intrusive, further increasing costs for banks and have a negative impact on banks' ability to innovate. It is against this background that the expansion of labour in the banking industry and subsequent decline in productivity can be associated with regulatory changes. The response of banks to regulatory changes in the form of expansion of labour are at the cost of reduced productivity in the sector. Put, otherwise, regulatory changes that fostered greater employment of labour that is not matched by an equal or greater expansion in bank output, particularly credit extension,[14] has translated into reduced labour productivity and constant unit cost of intermediation in the sector.

[11]Available at https://www.pwc.co.za/en/assets/pdf/major-banks-analysis-sept-2017.pdf.
[12]The report was prepared by the Department of Economic Development and Tourism and the University of Cape Town. Available at http://www.aifmrm.uct.ac.za/wp-content/uploads/AIFMRM-DEDT-Financial-Services-Sector-Assessment-Report-2014.pdf.
[13]The CEO's are quoted in article, available at https://www.businesslive.co.za/archive/2012-12-13-local-banks-decry-onslaught-of-new-rules/.
[14]The year on year growth in credit extension, which is a basic measure of bank output, peaks at just 2% post 2008 relative to a peak of 10% before 2008.

4 | CONCLUSION

This paper analyses bank productivity in South Africa and evaluates whether the development of the banking industry has translated into lower cost of intermediation or improved productivity and efficiency of the sector. Results from our computation shows that there is no apparent trend in the unit cost of intermediation between 1993 and 2019, which entails constant productivity. Therefore, despite major improvements in the sector with regards to technological advances, the cost of creating and maintaining intermediated assets has been constant. This can partly be associated with regulatory changes that occur post the GFC, which fostered an expansion of employment in the sector with the aim of implementing the regulatory changes. The expansion of employment is however not matched by an equal or greater expansion in bank output, which consequently translates into reduced labour productivity and constant unit cost of intermediation. Further research analysing the impact of prudential regulation in reducing the dynamism and efficiency of the sector is therefore suggested.

ORCID

Keaoleboga Mncube https://orcid.org/0000-0003-1652-3829
Nicola Viegi https://orcid.org/0000-0002-7181-6686

REFERENCES

Bazot, G. (2017) Financial consumption and the cost of finance: measuring fi- nancial efficiency in Europe (1950–2007). *Journal of the European Economic Association*, 16, 123–160. Available from: https://doi.org/10.1093/jeea/jvx008

Buera, F.J., Kaboski, J.P. & Shin, Y. (2011) Finance and development: a tale of two sectors. *American Economic Review*, 101(5), 1964–2002. Available from: https://doi.org/10.1257/aer.101.5.1964

Curdia, V. & Woodford, M. (2016) Credit frictions and optimal monetary policy. *Journal of Monetary Economics*, 84, 30–65. Available from: https://doi.org/10.1016/j.jmoneco.2016.10.003

Greenwood, J., Sanchez, J.M. & Wang, C. (2010) Financing development: the role of information costs. *American Economic Review*, 100(4), 1875–1891. Available from: https://doi.org/10.1257/aer.100.4.1875

Haldane, A., Brennan, S. & Madouros, V. (2010) What is the contribution of the financial sector: Miracle or mirage. In: Turner, A., Layard, R., Boone, P., Goodhart, C., Haldane, A., Johnson, S., *et al.* (Eds.) *The future of finance: the LSE report*. London School of Economics & Political Science, pp. 87–120.

Levine, R. (1997) Financial development and economic growth: views and agenda. *Journal of Economic Literature*, 35, 688–726.

Maredza, A. & Ikhide, S. (2013) Measuring the impact of the global financial crisis on efficiency and productivity of the banking system in South Africa. *Mediterranean Journal of Social Sciences*, 4, 553. Available from: https://doi.org/10.5901/mjss.2013.v4n6p553

Mlambo, K. & Ncube, M. (2011) Competition and efficiency in the banking sector in South Africa. *African Development Review*, 23(1), 4–15. Available from: https://doi.org/10.1111/j.1467-8268.2010.00268.x

Okeahalam, C.C. (2006) Production efficiency in the South African banking sector: a stochastic analysis. *International Review of Applied Economics*, 20(1), 103–123. Available from: https://doi.org/10.1080/02692170500362819

Philippon, T. (2015) Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *American Economic Review*, 105(4), 1408–1438. Available from: https://doi.org/10.1257/aer.20120578

Philippon, T. & Reshef, A. (2013) An international look at the growth of modern finance. *Journal of Economic Perspectives*, 27(2), 73–96. Available from: https://doi.org/10.1257/jep.27.2.73

Rashid, H., 2011. Report of the un-desa scoping mission to South Africa.

Verhoef, G. (2009) Concentration and competition: the changing landscape of the banking sector in South Africa 1970–2007. *South African Journal of Economic History*, 24(2), 157–197. Available from: https://doi.org/10.1080/10113430909511222

How to cite this article: Mncube, K. & Viegi, N. (2023) Measuring the unit cost of financial intermediation in South Africa: A measure of bank productivity. *South African Journal of Economics*, 1–10. Available from: https://doi.org/10.1111/saje.12364

APPENDIX A

TABLE A1 Data sources.

Variable	Definition	Source
Credit to private sector (Overdrafts, loans and advances)	Sum advances to financial corporate sector, non-fin corporate sector, unincorporated business enterprises of households, households and NPO's, factoring debtors & other loans & advances to above agents.	SARB Historical DI900 converted to BA900, 181, 187 & 188
Credit to households	Overdrafts, loans & advances to households	SARB Historical DI900 converted to BA900, 185 & 192
Credit to non-financial corporate sector	Overdrafts, loans & advances to non-financial corporate sector	SARB Historical DI900 converted to BA900, 183 & 190
Credit to public sector (government)	Overdrafts, loans & advances to public sector	SARB Historical DI900 converted to BA900, 171 (172–179)
Gross Domestic Product at Market, Constant 2010 prices, R millions: Seasonally adjusted Prices		Statistics South Africa (Stats SA) P0441
Gross Value Added (GVA)	It is a measure of the contribution to GDP made by the finance sector: (income earned by the relevant factors of production)	SARB Quarterly Bulletin (S110–137) and Quantec
Household deposits	Deposits by households denominated in Rand and foreign currency	SARB Historical DI900 converted to BA900, 27 & 35
Corporate deposits	Deposits by non-financial corporate sector (Rand and foreign currency)	SARB Historical DI900 converted to BA900, 25 & 37
Government deposits	Deposits by government (central, provincial, social security and local)	SARB Historical DI900 converted to BA900, 34, 6, 7, 10, 11 & 14
Number of employees	Number of employees, full time in economy and banking	Stats SA Quarterly Employment Statistics (QES) P0277 & Survey of Total Employment and Earnings, Household and Labour Surveys and Population Censuses
Gross earnings	Payments for ordinary-time, standard or agreed hours for all employees: Total sum of the earnings including performance and other bonuses	Stats SA Quarterly Employment Statistics (QES) P0277 &Survey of Total Employment and Earnings,Household and Labour Surveys and Population Censuses
Real per capita average earnings	Gross earnings divided by number of employees	Stats SA Quarterly Employment Statistics (QES) P0277 &Survey of Total Employment and Earnings,Household and Labour Surveys and Population Censuses
Labour remuneration	The amount paid to employees	Survey of Total Employment and Earnings,Household and Labour Surveys and Population Censuses
Remuneration per employee	It is equal to the total compensation of employees, divided by the number of employees	Survey of Total Employment and Earnings, Household and Labour Surveys and Population Censuses

(Continues)

TABLE A1 (Continued)

Variable	Definition	Source
Johannesburg Stock Market (JSE) market capitalization (shares)		SARB #KBP2024J
Public sector domestic marketable debt	Includes the financial debt instruments of all the institutions, at all levels of general government and in all resident public corporations	SARB #KBP4564J
Labour productivity index	Measured as output per unit of labour	Quantec