Revisiting Portuguese banks' efficiency and productivity

Nuno Ribeiro
Banco de Portugal

Inês Tavares
Banco de Portugal

October 2021

Abstract
The inefficient allocation of banks' resources is pointed out as a vulnerability of the banking system from a financial stability point of view, being its assessment of the outmost importance. In this work, the performance of the Portuguese banking system between 2012 and 2019 is assessed. Concretely, through the estimation of a translog cost frontier, total factor productivity growth is computed and decomposed into the effect of cost efficiency, technological progress and returns to scale. For this purpose, banks are assumed to choose the cost minimizing combination of labour, capital and interest bearing debt to produce loans and other earning assets. It is possible to conclude that the distance at which banks operated from the frontier remained constant throughout the period under consideration, suggesting that structural long-term factors play a bigger role for inefficiency than time-varying factors. Further, despite not being statistically significant, technological progress has been the main driver for the total factor productivity growth observed, particularly in the beginning of the period. Additionally, evidence for the existence of constant returns to scale during this period was found. (JEL: C23, D24, G21)

1. Introduction

The inefficient allocation of banks' resources is pointed out as a major drag on their profitability and, consequently, is considered a vulnerability from a financial stability point of view. In fact, even in case banks are monitoring risks adequately, low levels of profit generation capacity raise concerns regarding banks' ability to withstand possible future shocks, which might deteriorate their capital position and cause a reduction in lending to the economy. Therefore, besides allowing banks to provide a wide range of financial services, their smooth functioning is necessary for the intermediation of funds, which is crucial for both the financial system and the real economy. Assessing banks' performance is thus of the utmost importance, and it is even more relevant given the existing low interest rate environment and the expected increased materialization of credit risk due to the current pandemic crisis.

Acknowledgements: The authors would like to thank the editor, Pedro Duarte Neves, Ana Cristina Leal, Inês Drumond, Carlos Santos, Paula Antão and participants at internal seminars of the Financial Stability Department for their comments and suggestions which much improved this article. The analyses, opinions and conclusions expressed herein are the sole responsibility of the authors and do not necessarily reflect the opinions of Banco de Portugal or the Eurosystem.

E-mail: ngtribeiro@bportugal.pt; itavares@bportugal.pt

In this work, the performance of the Portuguese banking system between 2012 and 2019 is assessed, capturing its developments since the peak of the sovereign debt crisis. Concretely, this study intends to answer the following questions through the estimation of a cost function: (i) what is the level of banks' cost inefficiency - i.e. how far banks stand away from the best practices – and how does it compare with common accounting based measures?; (ii) what drives banks' marginal costs and margins?; and (iii) what is the level of total factor productivity growth in the period under analysis and how is it decomposed into the effect of cost efficiency, technological progress and returns to scale? By doing this, it is possible to better understand whether changes in total factor productivity were driven by catching up to the cost frontier, by shifts in the frontier itself or by movements along the frontier. The methodology is, therefore, particularly appealing because it allows for the estimation of these three effects within the same econometric framework.

The literature on the cost efficiency of Portuguese banks includes the work by Boucinha *et al.* (2013), Mendes and Rebelo (1999), Canhoto and Dermine (2003), Pinho (2001) and Pinho and Lima (2008). These studies cover the period between 1987 and 2006. Hence, existing literature is complemented with this work by extending the analysis to a more recent period. The approach taken is similar to that followed in the majority of these papers in that a translog cost function is estimated using Stochastic Frontier Analysis (SFA). In contrast, in Canhoto and Dermine (2003) a non-parametric frontier was estimated using Data Envelopment Analysis (DEA).

Even though all works but the one by Mendes and Rebelo (1999) found that banks' efficiency has increased over the period under analysis, they are not aligned on efficiency levels due to differences in the empirical and theoretical approaches to the modelling of banks' activity. Additionally, some studies do not allow for conclusions concerning technological progress, since the frontier is assumed to be constant over time. Nevertheless, while Mendes and Rebelo (1999) found evidence for the existence of technological recess in the period between 1990 and 1995, Boucinha *et al.* (2013) concluded that technological progress has shifted the cost frontier downwards throughout the period between 1992 and 2006, being the latter the main driver for the increase in productivity recorded in that period. Moreover, both the papers by Mendes and Rebelo (1999) and by Boucinha *et al.* (2013) concluded that scale economies contributed to increase banks' performance.

In what concerns the efficiency of European banks, Maudos *et al.* (2002) found that profit efficiency is lower than cost efficiency, considering a sample period between 1993 and 1996[1]. Interestingly, they also concluded that there is considerable variability in efficiency levels in the banking systems of the European Union, being this variability greater in terms of profit efficiency than in terms of cost efficiency. Nitoi and Spulbar (2015), in turn, used a heteroscedastic stochastic frontier model to investigate differences in cost efficiency of banks in six Central and East European countries over the period

1. A bank is the more cost (or profit) efficient, the closer its actual costs (or profits) are to the industry best practices.

from 2005 to 2011. They found that all banking systems in their sample recorded an increase in efficiency until 2008. However, they noticed that efficiency either stagnated or declined after 2009.

Focusing now on the identification of shifts in best practices, Altunbas *et al.* (1999) concluded that the rate of reduction in the costs of European banks due to technological progress increased between 1989 and 1996.

Huljak *et al.* (2019), in turn, noticed that, for euro area countries, total factor productivity increased between 2006 and 2017, even though at a decreasing pace. Technological progress was pointed out as the main driver for total factor productivity growth, followed by technical efficiency. Furthermore, they disentangled permanent and time-varying inefficiency and showed that the largest part of bank inefficiency in the euro area stems from persistent inefficiency. Lastly, they argued that, given the need to boost productivity and enhance profitability in the euro area banking sector, and since they found evidence for the existence of scale economies, possibly mergers and acquisitions should be intensified, along with banks' efforts in areas such as the rationalisation of branch networks and the digitalisation of business processes.

Finally, yet importantly, Oliveira (2017) uses SFA to characterize the production function of financial intermediation in Europe between 2000 and 2013 and concluded that, while there is ambiguous evidence on productivity growth, inefficiency of financial intermediation has been increasing over time, possibly driven by the least efficient banks. In addition, while increasing returns to scale were limited to smaller banks, scope savings were found to be robust across all models for the average bank. Interestingly, the main contribution of this work was to show that conclusions on the level of cost efficiency depend on the choice of the indicator for the bank-specific cost of funding.

This article proceeds as follows: Section 2 presents the methodology for the estimation of Portuguese banks' cost function and the data used. Section 3 contains the empirical results of this work and is divided into 6 subsections concerning the estimation results, marginal costs and margins, cost efficiency, technological progress, scale economies and total factor productivity change, respectively. Finally, Section 4 concludes.

2. Methodology and data

The literature aimed at studying banks' production typically differs in how deposits are modelled and in whether the financial structure is taken into account through the inclusion of equity.

Regarding the modelling of deposits, and according to the *intermediation approach*, the definition of cost includes deposit-based debt (with its price being part of the cost function) and, consequently, deposits are treated as inputs. Banks' main activity is, therefore, to grant loans and invest in securities and other assets using deposits and other funding as inputs, along with physical capital and labour[2]. Contrarily, the so-called *production approach* stresses the role of deposits in providing immediacy of transaction and payment services and the fact that physical resources are consumed in their production. Concretely, according to the latter approach, deposit-based debt would be treated as an output, being its quantity, rather than its price, included in the cost function.

In this work, following Boucinha *et al.* (2013), the intermediation approach is adopted, emphasising the role of deposits as sources of loanable funds in the intermediation process[3]. It should be mentioned that this approach allows for a more comprehensive definition of banks' costs, not limiting the measurement of efficiency to operational costs. In this way, the fact that some banks might be willing to bear higher operational costs (for example, with employees and equipment) in order to attain lower funding costs is taken into account.

Turning now to the question of the inclusion of equity as an input, most studies in the literature do not assign equity capital a role in the technology of intermediation and, consequently, the cost function fails to control for its level[4]. The reason for this is that it is difficult to obtain a price for this input. However, controlling for the price of other inputs, excluding equity could spuriously indicate that banks which rely more on equity are more efficient. Consequenlty, and in line with Boucinha *et al.* (2013), equity was tentatively included as a fixed input (being its price excluded from the cost function). Regulatory and rating/reputation issues, as well as the fixed costs associated to common equity issuances, may lead banks to have a higher level of equity than that yielded by the static cost minimization problem, justifying the treatment of equity as a fixed input. Nevertheless, evidence that a higher level of equity was associated with lower costs with the other inputs when controlling for the price of these inputs was not found. Along these lines, a negative value for the shadow cost of equity was obtained, which is not plausible since that would mean that bank shareholders would have to pay to hold equity[5]. Therefore, banks' financial structure was neglected by excluding equity from the analysis.

Banks are assumed to choose the amount of labour (L), funding (F) and physical capital (K) that minimizes the sum of its respective costs ($w_L L + w_F F + w_K K$) subject to the production of a given amount of loans (y_1) and other earning assets (y_2). Hence, the cost function $C(y_1, y_2, w_L, w_F, w_K)$ results from:

2. Capital is referred as "physical capital" for ease of exposition and to avoid confusion with the measure of capital in the funding structure, even though it includes intangibles, as well as tangible assets.

3. To motivate their choice, Boucinha et al. (2013) compute the elasticity of total costs deducted of interest paid on deposits to the level of deposits for the Portuguese banking system. If this elasticity is positive (negative), deposits should be modelled as outputs (inputs).

4. See Hughes *et al.* (2001) for more details.

5. The shadow cost of equity is given by the symmetric of the derivative of the cost function with respect to equity. Even though there are not many studies in the literature including equity in the cost function, Oliveira (2017) also obtained a negative value for this estimate for a sample of European banks.

$$\min_{L,F,K} \quad (w_L L + w_F F + w_K K) \tag{1}$$

$$s.t. \quad P(L,F,K) \quad \geq \quad y_1 + y_2, \tag{2}$$

where the price of labour (w_L) is computed as the ratio between labour costs and the average number of employees, the price of funding (w_F) is defined as the ratio between the flow of interest paid and the average stock of interest bearing liabilities, and the price of physical capital (w_K) is proxied by the ratio between the sum of depreciation and other general administrative costs (excluding labour) and the average stock of tangible and intangible assets[6].

It should be emphasized that loans and other earning assets are considered *net of impairments*. By doing so (instead of considering gross amounts), differences in both the level and the quality of banks' screening and monitoring in lending activities are taken into account, which are reflected in different intensity of materialisation of credit risk. Ultimately, impaired assets are a non-income producing item in banks' balance sheet and impairments measure the expected non-recoverable part of it. Nevertheless, it was not possible to take into account other differences in banks' management that might affect their efficiency. Concretely, while banks might target diverse loan segments (with, for example, different levels of monitoring costs and risk associated to mortgages, consumer loans, SME and large corporate loans), total (customer) loans were considered as a single homogeneous product. This might be a particularly relevant limitation since the heterogeneity of output has a bearing in the definition of the production possibility frontier. To overcome this problem, the output would have to be disaggregated into different categories of loans, which, in this case, is not feasible given the sample size and the required increase in the number of parameters to be estimated. However, it was not found evidence in the sample that investment banks and consumer credit banks are associated with differentiated cost efficiency levels. Further, both types of banks exhibit considerable dispersion around the overall sample mean.

In order to account for poor managerial performance (due to agency problems, for example), as well as for random factors that affect this performance, and for measurement error in the variables used in the estimation, each bank i's observed costs at time t can be written as:

$$C_{i,t} = C(y_{1,t}, y_{2,t}, w_{L,t}, w_{F,t}, w_{K,t}) \exp(v_{i,t}) \exp(u_{i,t}), \tag{3}$$

where $v_{i,t}$ is a random error assumed to follow an *i.i.d.* $N(0, \sigma_v^2)$ distribution that reflects both the effect of random uncontrollable shocks and measurement error. Thus, $C(y_{1,t}, y_{2,t}, w_{L,t}, w_{F,t}, w_{K,t}) \exp(v_{i,t})$ constitutes the stochastic frontier and $C(y_{1,t}, y_{2,t}, w_{L,t}, w_{F,t}, w_{K,t})$ the deterministic part of it. $u_{i,t}$, in turn, is a non-negative random variable which measure cost inefficiency. Following Battese and Coelli (1992), $u_{i,t}$ is defined as:

6. For factor prices calculation purposes, and for each year t, simple averages between the value of the relevant variable at the end of period t and the same value at the end of period $t-1$ are computed.

$$u_{i,t} = u_i \exp(-\eta(t - T_i)), \tag{4}$$

where u_i is assumed to follow an *i.i.d.* truncated normal distribution, $N^+(\mu, \sigma_u^2)$, T_i is the last available period for bank i and η is a decay factor. μ and η are parameters to be estimated. Given this specification, if η is statistically different from zero, inefficiency is considered to vary monotonically throughout time. Even so, and as a result of the specification in equation (4), the ranking of banks in terms of cost-inefficiency is preserved throughout the sample period. Otherwise, if η equals zero, inefficiency is considered to be time-invariant, and in that case, the parameter η is constrained to zero, so as to maximize the degrees of freedom in the estimation. It should also be mentioned that, according to the specification presented, inefficiency is firm specific.

Applying a natural log transformation to equation (3), the main equation to be estimated using SFA models may be expressed as:

$$\ln(C_{i,t}) = \ln(C(y_{1,t}, y_{2,t}, w_{L,t}, w_{F,t}, w_{K,t})) + v_{i,t} + u_{i,t}. \tag{5}$$

Once again, following Boucinha *et al.* (2013), the cost function is estimated using a translog functional form, which is a second order local approximation to the solution of the cost minimization problem for the average bank. This functional form is very popular in the literature, as it represents a balance between flexibility and parsimony. Indeed, while providing a good local approximation to the true cost function, it allows us to avoid multicolinearity problems and to preserve degrees of freedom, which is particularly relevant given the relatively small number of observations in the sample. Equation (5) for outputs r, s and inputs m, n can then be rewritten as:

$$\begin{aligned}
\ln(C_{i,t}) = {} & \gamma_0 + \gamma_t t + \frac{1}{2}\gamma_{t,t} t^2 + \sum_r \gamma_{t,r} \ln(y_{r,i,t}) t + \sum_m \gamma_{t,m} \ln(w_{m,i,t}) t \\
& + \sum_r \gamma_r \ln(y_{r,i,t}) + \sum_m \gamma_m \ln(w_{m,i,t}) + \frac{1}{2} \sum_r \sum_s \gamma_{r,s} \ln(y_{r,i,t}) \ln(y_{s,i,t}) \\
& + \frac{1}{2} \sum_m \sum_n \gamma_{m,n} \ln(w_{m,i,t}) \ln(w_{n,i,t}) + \sum_m \sum_r \gamma_{m,r} \ln(w_{m,i,t}) \ln(y_{r,i,t}) \\
& + v_{i,t} + u_{i,t},
\end{aligned} \tag{6}$$

with

$$
\begin{aligned}
&\gamma_{m,n} = \gamma_{n,m}, \quad \forall m, n \\
&\sum_m \gamma_m = 1 \\
&\sum_m \gamma_{m,r} = 0, \quad \forall r \\
&\sum_n \gamma_{m,n} = 0, \quad \forall m, n \\
&\sum_m \gamma_{t,m} = 0.
\end{aligned} \tag{7}
$$

It should be noted that the theoretical restrictions stemming from duality theory are imposed. Namely, symmetry is imposed as a result of the specification of the estimated equation, whereas linear homogeneity in prices is obtained by normalizing input prices and total cost by w_K.

Additionally, it should be mentioned that, in order to facilitate the interpretation of the parameters of the model, the data are expressed as deviations from the overall sample mean, so that the first order coefficients correspond to the elasticities evaluated at the sample mean.

The database comprises an unbalanced panel of yearly data on a consolidated basis including the major Portuguese banks between 2012 and 2019[7]. Concretely, the sample is composed of 15 banks and, for each year, it covers at least 68% of total loans, 66% of total assets and 76% of total deposits in the Portuguese banking system[8]. All data was retrieved from BankFocus by Moddy's Analytics and Bureau Van Dijk (BvD).

3. Empirical results

In this section, the main results of this work are presented. Concretely, through the estimation of a cost function, detailed in Subsection 3.1, banks' marginal costs and margins are studied in Subsection 3.2. Subsection 3.3 proceeds with the analysis of the evolution of a measure of cost efficiency. Subsections 3.4 and 3.5 examine the existence of technological progress and scale economies, respectively. Finally, in Subsection 3.6, total factor productivity change is computed and decomposed into the effect of cost efficiency, technological progress and returns to scale throughout the sample period.

---

7. The choice of consolidated accounts instead of solo basis unconsolidated data allows to include in the same economic unit all banks and other financial institutions belonging to the same banking group. Notwithstanding this advantage, both domestic and international activity are being considered, which are less likely to potentially share the cost base.

8. Novo Banco is excluded from the sample due to its recent strong restructuring process.

3.1. *Cost frontier estimation*

Table 1 presents cost frontier estimation results, underlying the estimation of equation (6) in Section 2. In accordance with what was mentioned in the previous section, since variables are expressed as deviation from the overall sample mean, it is possible to focus directly on single parameters to assess relevant elasticities at the sample mean. Hence, for simplicity, cross terms with no direct interpretation are omitted in the Table.

	(1)	(2)
$\ln(w_L)$	0.4367***	0.4070***
	0.0761	*0.0804*
$\ln(w_F)$	0.4220***	0.4434***
	0.0893	*0.0954*
$\ln(y_1)$	0.6326***	0.6259***
	0.0439	*0.0549*
$\ln(y_2)$	0.3163***	0.3247***
	0.0480	*0.0623*
$\ln(y_1)\ln(y_2)$	-0.0695*	-0.0849*
	0.0408	*0.0506*
t	-0.0112	0.0046
	0.0250	*0.0259*
η	-0.0404	
	0.0538	
Number of observations	82	82
Number of parameters	25	24
Log-likelihood	61.20	59.82
μ	0.1916	-1.1757
γ	0.8234	0.9700
σ^2	0.0464	0.2900
σ_u^2	0.0382	0.2813
σ_v^2	0.0082	0.0087

TABLE 1. Cost frontier estimation results

Notes: The constant and most cross terms were omitted.
Robust standard errors in italics.
$*p < 0.1, **p < 0.05, ***p < 0.01$

In the first regression presented in Table 1, the estimate for bank specific cost inefficiency is allowed to vary throughout time (see equation (4)). However, since η is not statistically significant, it is constrained to zero in regression (2). This being said, the analysis of results that follows is based on the specification underlying column (2).

It is worth to mention that, given the number of observations in the sample, the small number of degrees of freedom constitutes a limitation of this work, particularly when compared to Huljak *et al.* (2019), which deals with a much larger sample. However, it should be born in mind that the degrees of freedom in this work are of the same order of magnitude of the ones in previous studies for the Portuguese market cited in Section 1 (see for instance Boucinha *et al.* (2013)). Still related to this point, it was assessed whether it was advantageous to estimate such a large number of parameters using a translog functional form versus a Cobb-Douglas one - much simpler, yet much more restrictive. Concretely, a Wald test on the joint significance of the non-direct parameters of the translog cost function was performed and the latter were found to be significant at the 1% level.

In line with economic intuition, the elasticity of cost with respect to each of the input prices is positive. In addition, the sum of the estimates of the two outputs (y_1 and y_2) is close to one, suggesting the existence of close to constant returns to scale at the sample mean. This is confirmed by a formal test presented further ahead. It should also be emphasized that the value of the estimated parameter associated with loans is about twice the value of the estimated parameter associated with other earning assets, which indicates that providing additional loans is more resource consuming than it is to invest in other assets due to the screening and monitoring costs involved in granting loans. The coefficient of the interaction term between the two outputs, in turn, is negative, pointing to the existence of scope economies in the joint production of loans and other earning assets. Furthermore, there are not statistically significant technological changes affecting banks' cost structure at the sample mean, as can be inferred from the estimated parameter associated with the time trend. Finally, it is worth to mention that, through the analysis of the parameter γ, it is possible to infer that most of the total error's variance is accounted for by cost inefficiency rather than by the classical random error, reinforcing the importance of assessing banks' performance.

3.2. *Marginal costs and margins*

Bank specific marginal cost estimates for both the production of loans and the production of other earning assets can be obtained as follows:

$$mc_{r,i,t} = \frac{\partial C_{i,t}}{\partial y_{r,i,t}} = \frac{C_{i,t}}{y_{r,i,t}} \frac{\partial \ln(C_{i,t})}{\partial \ln(y_{r,i,t})}. \tag{8}$$

The first two time series presented in Table 2 are based on the estimated parameters for the cost function and are constructed by aggregating the individual estimates for marginal costs, using each bank's market share in loans as weights. The same weights are used to aggregate all the measures presented hereafter.

Given the decline in interest rates registered during the period under analysis (column 3) and the relevance of funding costs in banks' cost structure, the estimated decrease in marginal costs was already expected. However, it is relevant to understand whether there are other explanations for this decrease and, in particular, how non-financial marginal costs evolved during this period. For this purpose, so-called *real resource marginal costs* (columns 5 and 6) were computed, which are obtained by deducting the estimated marginal costs (columns 1 and 2) for each bank by the corresponding price of funding (column 4). As shown in the last two columns of Table 2, the latter are considerably more stable than total marginal costs, suggesting that the behaviour of interest rates is indeed the main driving factor for the fall in marginal costs (even though the real resource marginal cost of loans still exhibits a declining trend).

Year	Marginal cost of loans (1) (%)	Marginal cost of other earning assets (2) (%)	Short-term money market interest rate (3) (%)	Implicit price of funding (4) (%)	Real resource marginal cost of loans (5) (%)	Real resource marginal cost of other earning assets (6) (%)
2012	4.76	4.56	0.57	2.87	1.89	1.69
2013	4.22	4.39	0.22	2.40	1.81	1.99
2014	3.85	4.09	0.21	2.18	1.67	1.91
2015	3.03	3.12	-0.02	1.60	1.48	1.57
2016	2.62	3.05	-0.26	1.17	1.46	1.89
2017	2.43	2.99	-0.33	0.91	1.52	2.07
2018	2.15	2.60	-0.32	0.81	1.34	1.79
2019	1.85	2.13	-0.36	0.67	1.19	1.47

TABLE 2. Marginal cost estimates

Note: Total loans are used to compute weighted means.

Table 3 allows us to analyse the evolution of banks' price-cost margin on loans throughout the sample period. A measure for the latter was obtained by subtracting the marginal cost of loans from the implicit interest rate on loans, which was computed using data on banks' loan related interest income and stock of outstanding loans.

Year	Marginal cost of loans (1) (%)	Implicit interest rate on loans (2) (%)	Margin on loans (3) (%)	Cost of risk (4) (%)
2012	4.76	3.65	-1.10	1.61
2013	4.22	3.15	-1.06	1.12
2014	3.85	3.21	-0.64	1.42
2015	3.03	2.73	-0.30	0.87
2016	2.62	2.59	-0.03	1.88
2017	2.43	2.41	-0.02	0.38
2018	2.15	2.45	0.30	0.32
2019	1.85	2.42	0.57	0.23

TABLE 3. Margin on loans estimates

Note: Total loans are used to compute weighted means.

As shown in Table 3, the interest income on an additional loan was not enough to cover the associated cost during the majority of the period under analysis. In fact, in the wake of the sovereign debt crisis, the cost of funding, including retail deposits, was exacerbated, while the pricing of credit did not adjust immediately. Furthermore, several banks went through adjustment processes in the beginning of the sample period which eventually slowed down over time, incurring in higher costs in the first years. Contrarily, in the last year of the sample period, the interest income on loans was estimated to cover not only the funding cost, but also the cost of risk (column 4)[9].

3.3. *Cost efficiency*

The cost efficiency of a given bank can be defined as the ratio between the minimum cost it would operate with assuming no inefficiency exists (which is given by the stochastic frontier) and the cost level with which it operates:

9. The cost of risk is computed as the ratio between loan impairments and the amount of outstanding loans.

$$CE_{i,t} = \frac{E[C|u_{i,t} = 0, X_{i,t}]}{E[C|u_{i,t}, X_{i,t}]} = \frac{C(y_{1,t}, y_{2,t}, w_{L,t}, w_{F,t}, w_{K,t}) \exp(v_{i,t})}{C_{i,t}} \in (0,1), \quad (9)$$

where $X_{i,t}$ are the regressors underlying column (2) of Table 1 (see Subsection 3.1). It is worth to refer that this measure lies between 0 and 100%, being higher values associated with higher efficiency levels. Intuitively, a fully efficient bank would have an efficiency level of 100%, indicating that its actual cost is exactly on the cost frontier.

Table 4 presents aggregate estimates for the cost efficiency measure between 2012 and 2019, and also for the entire period.

Year	CE (%)
2012	83.81
2013	83.22
2014	83.31
2015	84.35
2016	84.38
2017	85.17
2018	84.56
2019	84.47
2012-2019	**84.10**

TABLE 4. Cost efficiency
Note: Total loans are used to compute weighted means.

The overall estimate for cost efficiency in this period stands at 84%, suggesting that Portuguese banks could have produced the same output while incurring only 84% of their actual costs[10]. Some heterogeneity was found across banks, with estimated cost efficiency ranging from 66% to 98% in the period under analysis. However, it is possible to infer that the distance at which banks stand from the cost frontier representing best practices did not change significantly over time, which suggests that structural long-term factors (for example, location, client structure, macroeconomic development and, to a lesser extent, regulation) play a bigger role for inefficiency than time-varying factors[11]. This is consistent with the final (time invariant) specification for the model.

In order to understand to which extent this measure of cost efficiency adheres to other common accounting based measures, Figure 1 illustrates the relationship between the former, on the one hand, and operational cost to core income (CCI) and operational cost to total assets (CA), on the other, for all the banks in the sample in 2019[12].

According to what was expected, the cost efficiency measure presented in this work is negatively correlated with these accounting measures, usually computed to assess the

10. Similar values were found in previous studies with comparable methodologies, namely 83% in Boucinha *et al.* (2013) for the Portuguese market in the 1990-2006 period, and 84% in Huljak *et al.* (2019) for the median bank in the euro area in the 2006-2017 period.

11. Huljak *et al.* (2019) reach the same type of conclusion by disentangling time-varying and persistent inefficiency.

12. Core income was computed as the sum between net interest income and net fees and commissions. Core income (rather than total income) was considered to abstract from extraordinary financial gains/losses.

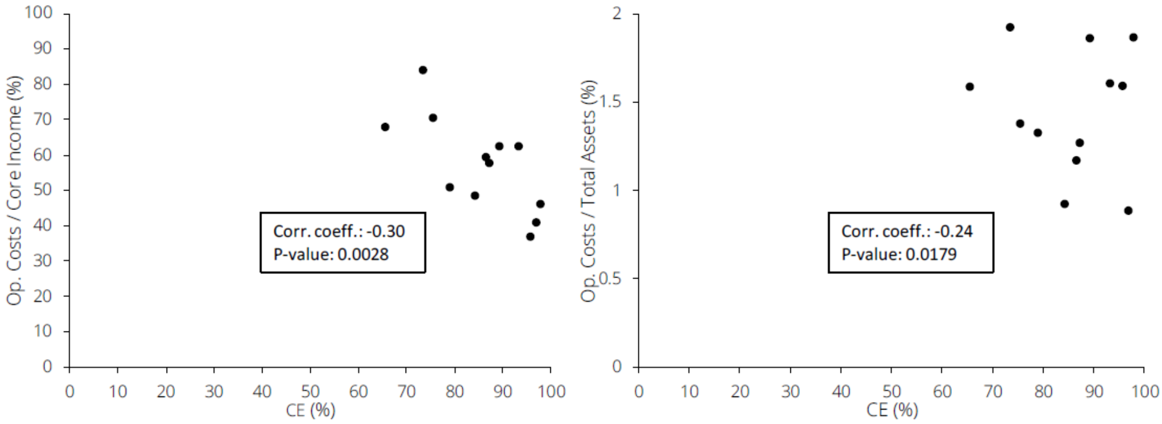

FIGURE 1: Cost efficiency vs other common efficiency measures

quality of banks' management. Concretely, the correlation coefficient between these two measures and the cost efficiency measure stands at -0.30 (statistically significant at the 5% level) and -0.24 (statistically significant at the 10% level), respectively. It should be born in mind, however, that these measures consider only operational costs, while cost efficiency includes a more comprehensive definition of banks' costs.

Despite being common to use accounting indicators to proxy efficiency in the banking sector, it is worth to refer that, as pointed by Huljak *et al.* (2019), these indicators have considerable shortcomings. While the CA ratio is strongly dependent on the business model of the institutions and their size, the CCI ratio is highly dependent on the income component, which, besides being affected by the degree of market power, is also influenced by credit risk, further distorting the measurement of efficiency. In fact, a bank that grants loans to a riskier counterpart will demand a higher interest rate than another which is equally efficient but more risk-averse[13].

3.4. Technological progress

In this subsection, the existence of technological progress - *i.e.*, shifts in the cost frontier brought about by the adoption of more efficient production techniques – is examined. The latter are given by the symmetric of the semi-elasticity of total cost with respect to the time trend[14]:

$$TP_t = -\frac{\partial \ln(C)}{\partial t}, \tag{10}$$

13. Even though there is a positive relationship between the rate charged on loans and the one paid on interest-bearing liabilities, which could attenuate the distortion from higher risk-taking, that relationship is neither one-to-one nor constant throughout time.

14. This measure underestimates technological progress when the quality/variety of products increases throughout time. This might be relevant in the sample period, notably due to the increased customization in the pricing of credit, which results from banks' more sophisticated models for credit risk assessment.

being positive values associated with technological progress, while negative values characterize a period of technological regression.

Table 5 presents estimates for technological change between 2012 and 2019, and also for the entire period. Interpreting the value for 2012, for example, it is possible to say that, in this year, Portuguese banks operating according to the industry's best practices could produce the same output as in 2011 incurring 2.9% lower costs. Even though total cost reducing technological progress was not statistically significant during the period under analysis, the rationalization of branch networks (which are more intense in both physical capital and labour inputs) seems to have been fruitful in the first half of the sample period, despite banks' simultaneous investment in back office activities, for instance in credit risk assessment and compliance domains. Tentatively, the latter factor may be behind the apparent slight technological recess in the latter years of the sample, as stricter and more comprehensive regulatory demands represent a cost increasing change in the environment in which banks operate. However, the resulting reputational benefits stemming from avoiding financial instability justify the incurrence in these costs.

Year	Technological progress (1) (%)	P-value (H0: $TP = 0$) (2)
2012	2.92	0.56
2013	2.09	0.64
2014	0.73	0.86
2015	0.96	0.81
2016	-0.53	0.89
2017	-0.69	0.85
2018	-1.16	0.75
2019	-1.30	0.77
2012-2019	**0.41**	**0.78**

TABLE 5. Technological progress

Note: Total loans are used to compute weighted means.

3.5. *Scale economies*

This subsection proceeds by assessing the presence of scale economies, which is a particularly relevant topic since it allows us to infer on the adequacy of the market structure from a technological point of view. A measure of scale economies in a multiproduct firm is obtained by summing up the elasticities of total cost with respect to each output:

$$SE_{i,t} = \sum_r \frac{\partial \ln(C_{i,t})}{\partial \ln(y_{r,i,t})}. \tag{11}$$

In the presence of scale diseconomies, the value for this measure is higher than one, while it is lower than one in the presence of scale economies.

As shown in Table 6, the period under analysis is characterized by constant returns to scale in statistical terms. However, numerical results point to the existence of scale diseconomies in the first half of the period, implying that costs would change more than proportionally to banks' size. In this case, deleveraging processes turn out to be advantageous from a cost reducing perspective, as shown in the next subsection.

Year	Scale economies (1)	P-value (H0: $SE = 1$) (2)
2012	1.07	0.65
2013	1.05	0.71
2014	1.04	0.76
2015	1.02	0.91
2016	1.01	0.96
2017	0.99	0.93
2018	0.97	0.80
2019	0.95	0.68
2012-2019	**1.01**	**0.90**

TABLE 6. Scale economies

Note: Total loans are used to compute weighted means.

It should born in mind, however, that the definition of cost underlying this analysis does not include the cost of equity. Thus, the measure of scale economies presented above is actually a measure of cash flow cost economies, which is likely to overestimate the true scale parameter. In fact, any increase in output must be exclusively financed by interest bearing debt, implying that total costs might be forced to increase more than would be necessary if banks could also use equity.

3.6. Total factor productivity change

In this subsection, the results presented in Subsections 3.3, 3.4 and 3.5 are brought together by decomposing total factor productivity (TFP) change into the contribution of cost efficiency change, technological progress and returns to scale, which correspond to catching up to the cost frontier, shifts in the frontier itself over time and movements along the frontier, respectively.

The decomposition that follows is borrowed directly from Bauer (1990), being its terms associated with previously mentioned effects:

$$
\begin{aligned}
\ln(\frac{TFP_{i,t}}{TFP_{i,t-1}}) &= \ln(\frac{CE_{i,t}}{CE_{i,t-1}}) \\
&+ \frac{1}{2}(-\frac{\partial \ln(C_{i,t})}{\partial t} - \frac{\partial \ln(C_{i,t-1})}{\partial t}) \\
&+ \frac{1}{2}\sum_r [(\varepsilon_{r,i,t}\frac{1-SE_{i,t}}{SE_{i,t}} + \varepsilon_{r,i,t-1}\frac{1-SE_{i,t-1}}{SE_{i,t-1}})\ln(\frac{y_{r,i,t}}{y_{r,i,t-1}})].
\end{aligned}
\tag{12}
$$

Table 7 summarizes the results for total factor productivity change during the period under consideration.

Year	Cost efficiency change (1) (pp)	Technological progress (2) (pp)	Returns to scale (3) (pp)	Total factor productivity change (4) (%)
2013	0.04	2.50	1.03	3.58
2014	0.00	1.41	0.33	1.74
2015	1.24	0.85	-0.21	1.88
2016	0.04	0.22	0.14	0.40
2017	0.95	-0.61	-0.01	0.33
2018	-0.72	-0.92	-0.16	-1.80
2019	-0.10	-1.23	0.06	-1.27

TABLE 7. Total factor productivity change decomposition

Note: Total loans are used to compute weighted means. To compute the technical efficiency change, a constant sample was considered for every two adjacent years.

Total factor productivity increased 4.9% between 2012 and 2019, with relevant positive changes in the initial years of the sample. Nevertheless, total factor productivity growth slowed down throughout the period, reaching negative values in the last two years.

Despite not being statistically significant, technological progress has been the main driver for the total factor productivity growth registered, particularly in the beginning of the period, with an annual average of 0.3 pp over the period comprised between 2011 and 2019[15]. Furthermore, even though the sample period is characterized by constant returns to scale in statistical terms, changes in scale efficiency also contributed positively to total factor productivity change in 2013 due to the existence of scale diseconomies, in conjunction with the strong deleveraging registered after the sovereign debt crisis. Throughout the period under analysis, scale efficiency changes contributed, on average, 0.2 pp per year to total factor productivity change[16]. In turn, the contribution of variations in cost efficiency to the annual total factor productivity change amounted to 0.2 pp, on average[17].

4. Concluding remarks

The inefficient allocation of banks' resources is pointed out as a vulnerability of the banking system from a financial stability point of view, being its assessment of the outmost importance. In this regard, the performance of the Portuguese banking system

15. Boucinha *et al.* (2013) also point technological progress as the most relevant factor driving total factor productivity growth for the period between 1992 and 2006, standing, on average, at 2.2 pp a year. The latter result is stronger than the one presented in this work and should be read against a background of intense liberalization, consolidation and privatization. On its turn, Huljak *et al.* (2019) estimate that the annual rate of technological progress amounted to 2.4 pp over the period from 2006 to 2017.

16. Once again, comparing with two previous studies that apply similar methodology, Boucinha *et al.* (2013) found the scale effect to equal 1.5 pp, on average, for the period from 1992 to 2006 (which included significant consolidation in the Portuguese banking system), while in Huljak *et al.* (2019) this estimate for euro area banks stood at around 0.25 pp, on average, during the 2006-2017 period.

17. Cost efficiency remained virtually unchanged in Boucinha *et al.* (2013) and its contribution to total factor productivity growth in the euro area was negative in the 2006-2017 period (and decreased from -0.8 pp in 2006 to -1.95 pp in 2017), as reported in Huljak *et al.* (2019).

between 2012 and 2019 is assessed, capturing its developments since the peak of the sovereign debt crisis. Concretely, through the estimation of a cost function, banks' marginal costs and margins are examined. Additionally, total factor productivity growth is computed and decomposed into the effect of cost efficiency, technological progress and returns to scale. The methodology is, therefore, particularly appealing because it allows for the estimation of these three effects within the same econometric framework. It is worth to refer, however, that, given the number of observations in the sample, the small number of degrees of freedom constitutes a limitation of this work. For the purpose of the analysis described above, banks are assumed to choose the cost minimizing combination of labour, capital and interest bearing debt to produce loans and other earning assets. This being said, total (customer) loans were considered as a single homogeneous product, which is a drawback since the heterogeneity of output has a bearing in the definition of the production possibility frontier.

Portuguese banks' marginal costs in the production of loans and other earning assets decreased during the period under consideration. In turn, real resource marginal costs, which are obtained by deducting the estimated marginal costs for each bank by the corresponding price of funding, were found to be considerably more stable, suggesting that the behaviour of interest rates is indeed the main driving factor for the fall in marginal costs. Furthermore, and contrarily to what happened in the latter years under analysis, the interest income on an additional loan was not enough to cover the associated cost during the majority of the sample period. This is justified, on the one hand, by the adjustment processes observed in the beginning of the period and, on the other hand, by the exacerbated cost of funding recorded in the wake of the sovereign debt crisis, together with the absence of sufficient adjustment in the pricing of credit.

The aggregate estimate for cost efficiency in the sample period lies close to 84%, suggesting that Portuguese banks could have produced the same output while incurring only 84% of their actual costs. This estimate, which reflects the distance at which banks stand from the cost frontier representing best practices, did not change significantly over time, suggesting that structural long-term factors (for example, location, client structure, macroeconomic development and, to a lesser extent, regulation) play a bigger role for inefficiency than time-varying factors. Therefore, structural policies that improve time-invariant efficiency of the Portuguese banking system should be considered, such as policies promoting digitalization in the economy. Concerning cost reducing technological progress, it was not statistically significant throughout the period. Moreover, it was not found evidence for the existence of increasing returns to scale in statistical terms, implying that scale increases through acquisitions do not seem to be advantageous.

Finally, total factor productivity increased 4.9% between 2012 and 2019, registering relevant positive changes in the first years. However, total factor productivity growth slowed down throughout the period, reaching negative values in the last two years. Despite not being statistically significant during the period under analysis, technological progress has been the main driver for the total factor productivity growth observed, particularly in the beginning of the period. Furthermore, changes in scale efficiency also contributed positively to total factor productivity change in 2013 due to the existence of scale diseconomies (although not statistically significant), in conjunction with the strong deleveraging registered after the sovereign debt crisis.

Overall, similarly to what was found for banks in other European countries Huljak *et al.* (2019), Portuguese banks have room to improve in what concerns their cost efficiency. That said, the efforts that banks have made in areas such as the rationalisation of branch networks and the digitalization of business processes are crucial to boost productivity and enhance profitability in the near future.

References

Altunbas, Yener, John Goddard, and Phillip Molyneux (1999). "Technical Change in Banking." *Economics Letters*, 64(2), 215–221.

Battese, G. E. and T. J. Coelli (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." *Journal of Productivity Analysis*, 3, 153–169.

Bauer, Paul W. (1990). "Decomposing TFP Growth in the Presence of Cost Inefficiency, Nonconstant Returns to Scale, and Technological Progress." *Journal of Productivity Analysis*, 1, 287–299.

Boucinha, Miguel, Nuno Ribeiro, and Thomas Weyman-Jones (2013). "An Assessment of Portuguese Banks' Efficiency and Productivity Towards Euro Area Participation." *Journal of Productivity Analysis*, 39, 177–190.

Canhoto, Ana and Jean Dermine (2003). "A Note on Banking Efficiency in Portugal, New vs. Old Banks." *Journal of Banking and Finance*, 27, 2087–2098.

Hughes, Joseph P., Loretta J. Mester, and Choon-Geol Moon (2001). "Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-taking Into Models of Bank Production." *Journal of Banking and Finance*, 25(12), 2169–2208.

Huljak, Ivan, Reiner Martin, and Diego Moccero (2019). "The Cost-Efficiency and Productivity Growth of Euro Area Banks." Working paper, European Central Bank.

Maudos, Joaquin, José Pastor, Francisco Perez, and Javier Quesada (2002). "Cost and Profit Efficiency in European Banks." *Journal of International Financial Markets, Institutions and Money*, 12(1), 33–58.

Mendes, Victor and João Rebelo (1999). "Productive Efficiency, Technological Change and Productivity in Portuguese banking." *Applied Financial Economics*, 9, 513–521.

Nitoi, Mihai and Cristi Spulbar (2015). "An Examination of Banks' Cost Efficiency in Central and Eastern Europe." *Procedia Economics and Finance*, 22, 544–551.

Oliveira, João (2017). "Inefficiency Distribution of the European Banking System." Working paper, Banco de Portugal.

Pinho, Paulo (2001). "Using Accounting Data to Measure Efficiency in Banking: an Application to Portugal." *Applied Financial Economics*, 11, 527–538.

Pinho, Paulo and Filipa Lima (2008). "Financial Disintermediation and the Measurement of Efficiency in Banking: the Case of Portuguese Banks." *International Journal of Banking, Accounting and Finance*, 1(2), 133–148.