TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

Bushra Faroquei

EFFICIENCY ANALYSIS OF EUROPEAN BANKS:

A TWO-STAGE DEA APPROACH

Bachelor's thesis

Programme International Business Administration, specialisation finance

Supervisor: Simona Ferraro, PhD

Tallinn 2022

I hereby declare that I have compiled the thesis independently and all works, important standpoints

and data by other authors have been properly referenced and the same paper has not been

previously presented for grading.

The document length is 11625 words from the introduction to the end of conclusion.

Bushra Faroquei

Student code: 194469TVTB

Student e-mail address: iibushrafaroquei9804@gmail.com

……………………………………………

Supervisor: Simona Ferraro, PhD

The paper conforms to requirements in force

……………………………………………

Chairman of the Defence Committee:

Permitted to the defence

…………………………………………

TABLE OF CONTENT

ABSTRACT

The aim of this thesis is to investigate the efficiency of banks from the European Union (EU) using an external two-stage Data Envelopment Analysis (DEA) technique. In total, 3952 observations from 654 banks during 2010-2020 are identified as the final sample. The selected banks are from 9 EU countries. The first stage of the analysis consists of two DEA models, namely the Banker, Charnes and Cooper (BCC) and Charnes, Cooper and Rhodes (CCR). The CCR model calculates technical efficiency avoiding the impact of the scale inefficiences, while the BCC model calculates only the pure technical efficiency scores. Therefore, adopting both CCR and BCC models allows measuring the scale efficiencies of banks.

The results of the DEA analysis suggest that EU banks, on average, have low technical efficiency scores, although they seem to operate quite closer to their most optimal production size as indicated by higher scale efficiency scores. Banks from Poland, Denamark, Italy, and Luxembourg have higher technical efficiency scores. Moreover, the findings of the thesis also suggests that smaller EU banks are more technically efficienct. Finally, the Tobit regression results found from the second stage of the analysis provides evidence that bank size and loans to asset ratio has a significant and negative impact on the EU bank efficiency in this thesis, while macroeconomic variable, per capita gross domestic product, has no significant association.

Keywords: Data Envelopment Analysis, bank efficiency, regression analysis, EU banking

INTRODUCTION

The topic of this thesis lies in measuring the bank efficiency and its determinants in the European Union (EU) banking sector. Banks are financial institutions that play a crucial role in a country's economic development and growth. As indicated by Wang et al. (2013), the banking sector is a highly complex industry that not plays an increasingly critical role in the development of the financial system and the economic growth of a country, but also impacts the day-to-day lives of the general public. Therefore, the evaluation of bank performance is important as it provides an indication about a bank's financial well-being to its different stakeholders, such as regulators, customers, and investors (Henrique et al. 2022). Besides profitability, which is also vital in a bank's long-term stability (Athanasoglou et al. 2008), efficiency is what essentially strengthens and guarantees the survival of banks.

Efficiency, as a concept, involves choosing a certain volume and structure of inputs and outputs which allows minimizing the cost or maximizing the profit of a firm or an institution. In simple words, efficiency can be thought of as a bank's ability to achieve its maximum level of output from a given level of input. Thus, efficiency measures the effectiveness of a firm to produce the minimum waste of time, effort, and skill (Sami et al. 2019). Although there have been several research conducted in the past on the topic of bank efficiency in countries all over the world, studies done on the banking sector of the European Union (EU) that includes banks from more than one country are still limited.

Besides overcoming the consequences brought forth by the recent global financial crisis, banks from majority of the EU countries also had to deal with challenges regarding the European integration, sovereign debt crisis, and the crises related to the European Economic and Monetary Union (EMU) (Ferreira 2019). Moreover, the recent changes in the EU regulations intend to push the industry in the direction of a single market, so the member nations should also have comparable levels of economic performance, especially in the banking system (Neves et al. 2020). This makes it even more interesting to study the bank efficiency in the different EU countries to see whether the efficiency levels are actually similar or vary a lot. Additionally, the factors determining various levels of bank efficiency are also important to study as it can help managers to take better decisions in improving bank performance (Neves et al. 2020). Therefore, the research problem of this paper is to not only investigate the level of EU bank efficiency, but also to find out the factors that caused the various levels of efficiency outcomes.

Data Envelopment analysis (DEA) is a non-parametric linear programming method that is widely used in studies to analyse bank efficiency. However, the traditional, or the simple DEA model is limited to only measuring efficiency of an organization, and do not reveal anything about the external factors that might have influenced the level of efficiencies. Two-stage DEA models have been introduced to overcome this issue, and it does so by including an additional stage of analysis to find out the efficiency determinants. So overall, the two-stage approaches can not only measure the EU banks' efficiency scores, but also investigate the impact of external factor on the estimated scores.

Therefore, the research aim of thesis is to to investigate the bank efficiency and its determinants for banks in the EU countries over the 11 year period of 2010-2020 using a two-stage DEA analysis. The thesis will attempt to answer the following research questions:

- How efficient is the banking sector in the selected European countries?
- What are the determinants that influence different levels of bank efficiency?

The structure of the thesis is split into three parts. The theoretical background section will consist of an overview of the various efficiency analysis techniques used in the previous studies, as well as a review of the empirical studies on bank efficiency conducted from both EU and non-EU countries. The data and methods section will present the selection of variables for the two stages of this thesis, some basic descriptive statistics to understand the characteristics of the selected dataset, and finally the model specific equations will be presented in the research methodology section. Finally, the findings and discussions section will present the results of the analysis, as well as comparing the results with those found by the previous empirical studies. The thesis will conclude with a summary of the main findings and some of the research limitations.

1. THEORETICAL FRAMEWORK

This part of the thesis is divided into two sections. The first section will provide an overview of some of the most popular efficiency analysis techniques used in the banking literatures. Followed by this, section 1.2 will include a broad discussion and review of previous empirical studies conducted to measure efficiency of banks using DEA approach, which is the chosen method for this thesis.

1.1 Efficiency analysis techniques

Studies have long ago proven that financial institutions like banks can be classified as decision making units (DMUs), where the best performing DMU is the one that operates on the efficient frontier, therefore acts as the best practice benchmark to allow an institutional ranking based on efficiency scores (Anagnostopoulos et al. 2022). There is no doubt that banks play a crucial role in a country's economic development and growth, and that effective evaluation of bank performance is highly needed to maintain their financial stability, especially in today's increasingly competitive market. Besides profitability, which helps a bank with long-term stability (Athanasoglou et al. 2008), efficiency is what essentially strengthens and guarantees the survival of banks.

The first approach to efficiency estimation was proposed by M.J.Farrell in 1957. It introduced the term ‘‘Frontier efficiency analysis’’, which is a tool that can be used to measure the productive efficiency of an industry. Figure 1 presents the concept of efficiency following the Farrell (1957) model where the assumption is that a firm is producing a single unit of product (output) by employing two factors of production (inputs), under the conditions of constant returns to scale. As explained by Trujillo and González (2008), the function of the efficient production can be given as, $y = f(x_1, x_2)$, where x_1 and x_2 represents the two inputs used in obtaining output y. With the constant returns to scale assumptions, the function for the SS'curve can be written as a unit isoquant $1 = f(x_1/y, x_2/y)$, which represents the different combinations of x_1 and x_2 used by a perfectly efficient firm to produce a unit of output (Farrell 1957). Now assuming Point P represents a firm that uses inputs in quantities x_1^* and x_2^* to obtain a unit of output y^*, and Point Q represents an efficient firm that uses the same proportion of inputs as P. Therefore, technical efficiency score for P can be measured as the ratio, OQ/OP, because this is the proportion of inputs that is absolutely necessary for P to produce the same quantity of outputs. This ratio, or the technical efficiency score, ranges between 1 and 0, where a value of 1 represents an efficient firm, while 0 represents an inefficient firm (Trujillo and González, 2008).

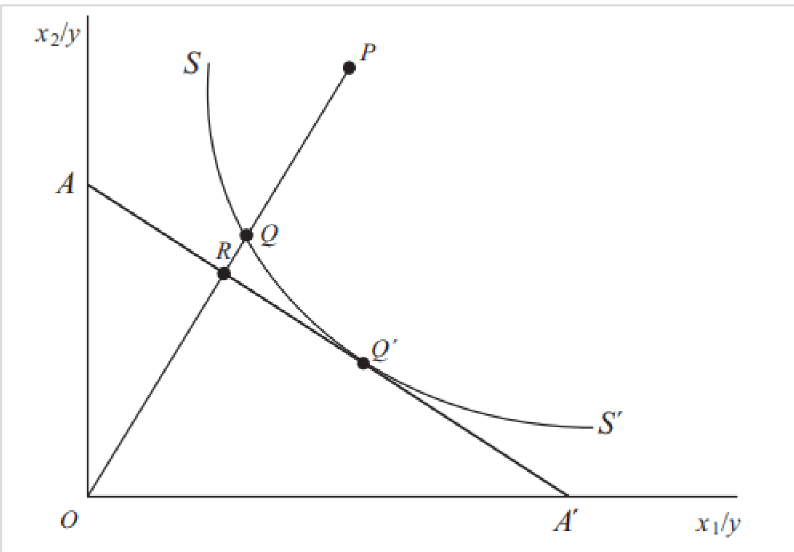

Figure 1: Technical and Allocative Efficiency (Farrell, 1957)
Source: Trujillo and González (2008)

Overall, Farrell's model showed that the frontier production function can be a relevant comparison for measuring productive efficiency (Førsund 2015). However, it was not until 1978 when Charnes, Cooper and Rhodes (CCR) introduced the term Data Envelopment Analysis (DEA) in their paper that efficiency analyses gained a broad popularity among the finance literatures. DEA is a non-parametric linear programming technique that is used to measure the productivity and efficiency of DMU's, and its primary goal is to improve the outputs of the DMUs, given the available inputs. While Farrell's model considered all DMUs as one identical production function and used the concept of multiple inputs and a single output, the CCR model expanded these specifications to using multiple inputs and multiple outputs, as well as measuring the relative efficiency of each DMU (Lin et al. 2009). The efficiency calculated is between 0 and 1, with 1 indicating a DMU is efficient, and 0 indicating that it is inefficient. The measurements are based on constant return to scale (CRS) assumptions in the CCR model (Charnes et al. 1978).

Later in the year 1984, Banker, Charnes and Cooper (BCC) further expanded both the Farrell and the CCR model, from efficiency being measured with constant returns to scale to variable returns to scale or VRS. The main difference between the two models is the treatment of the scale effects. The BCC model demonstrated that the CRS measure of technical efficiency can be expressed as the product of two components: the pure technical efficiency measure and the scale efficiency measure (Anagnostopoulos et al. 2022). Under the VRS assumptions, the BCC model calculates only the pure technical efficiency (PTE) scores, which considers only the pure resource-conversion efficiencies attained by DMUs, irrespective of whether they enjoy increasing, decreasing or constant returns to scale (Majumder, Chang 1996). Scale efficiency (SE), which is the second component of the CCR model, determines how close an observed DMU is to its most productive scale size. SE score ranges between 0 and 1, with 1 being scale efficient, and 0 being otherwise. The CCR model calculates the technical efficiency with the assumption that all the DMUs are operating at its optimal size, with SE=1, thus not considering the scale inefficiencies. But this is not a realistic scenario, as there exists scale inefficiencies when DMUs are not operating closest to their productive scale size, which tend to increase the average costs of production (Majumder, Chang 1996).

According to Banker et al (1984), scale efficiency can be calculated by dividing the CRS measure of TE under the CCR model ($\mathrm{TE}_{\mathrm{CRS}}$), over the VRS measure of PTE under the BCC model ($\mathrm{TE}_{\mathrm{VRS}}$). Thus, the equation can be written as follows:

$$SE = \frac{TE_{CRS}}{TE_{VRS}} \qquad (1)$$

Among various techniques and approaches used to estimate efficiency of financial institutions, DEA is quite often used in studies to measure bank efficiency. Despite having major similarities with the concept of efficiency in microeconomics, DEA does not require explicit assumption or specification of functional forms of the frontier to determine the production frontier. Instead, it is generated from the actual data for the analyzed firms (Casu et al. 2003). DEA also does not consider the existence of random errors in the measurements, which is considered as one of the limitations of this approach (Berger, Humphrey 1997). In other words, the efficiency score for a particular firm is defined relative to the other firms in the specific data set under consideration rather than by an absolute standard.

Free Disposal Hull (FDH) is another non-parametric approach, and it is a special case of the DEA method. It shares similar characteristics with DEA, i.e, allowing efficiency to vary over time and requiring no prior assumption about the kind of the inefficiency distribution across the observations (Berger et al. 1997). However, FDH differs by some key points, including that it usually results in higher average efficiency scores than the DEA estimations (Tulkens 1993).

Besides the non-parametric approaches mentioned above, parametric techniques such as the stochastic frontier approach (SFA) is also a popular method used in the studies dealing with banking efficiency. SFA estimates a parametric frontier of the best possible practices using a standard cost or profit function (Silva et al. 2017). The main assumption of this technique is that since inefficiencies cannot be negative, they must have an asymmetric or truncated distribution. This half normal assumption for inefficiency distribution can often give rise to separability issues with the symmetrically distributed random errors, which is one of the limitations when using this approach (Silva et al. 2017).

Parametric and non-parametric methodologies have different degrees of dispersion and rank banks differently, which is the core reason behind the conflict on which method is the best to use (Berger et al. 1997). For example, bank efficiency scores obtained by SFA and DEA approach may provide more or less similar conclusions at the global level but slightly vary at the individual level, causing rank dissimilarity (Silva et al. 2017). One of the reasons behind the difference in results could be due to DEA's inability to take random shocks into consideration. Nevertheless, both parametric and non-parametric approaches have a few limitations. While the first restricts the shape of the frontier by specifying the mathematical expressions and setting strict assumptions, the latter needs to adopt a set of assumptions about the error and efficiency terms (Berger et al. 1997).

Despite all the criticisms, DEA remains the most popular non-parametric approach in bank efficiency studies (Berger et al. 1997). It is widely used by researchers, and sometimes, even preferred over the parametric approaches because it is based upon a set of few assumptions and is more data driven (Esteve et al. 2022). DEA's main advantage lies in the fact that it is non-parametric, therefore can be easily computed using linear programming without requiring any functional form specifications (Strange et al. 2021). Moreover, DEA can deal with the challenge of analyzing multidimensional situations that includes setting multiple inputs against multiple outputs multi-output situations in an easier way (Aggelopoulos, Georgopoulos 2017, pp 1172).

Although traditional DEA models have maintained their popularity throughout the years, numerous studies have shed light on some important issues, or rather biases, associated with their efficiency estimations, which may require improvements. Firstly, the simple DEA approach does not reveal anything about the transformation structure within the DMUs production process that generates the outputs using the inputs. In other words, it is as if the transformation is taking place inside a so-called black box, without being impacted at all by the DMU's external or environmental factors (Henrique et al. 2022). While this lack of too much structure is indeed one of the benefits of using DEA as discussed above, it is also a limitation for complex industries like the banking sector which requires a more structured model. To overcome these limitations, two-stage DEA models were introduced, which have already gained prominence among the financial literatures, especially the bank efficiency studies.

Two-stage DEA models can be either internal or external. Internal two-stage models divide the production process into two subprocesses, involving the use of intermediate variables within the production process. On the other hand, external models involve applying some techniques, typically a regression analysis, where the dependent variables are the efficiency scores estimated through the DEA in the first stage. Hence, the second stage of the analysis is taking place outside of the production process, unlike the internal models (Henrique et al. 2022). While both models can resolve the black box problem, the application of another technique after DEA (external) may allow for a more complete analysis for banks by taking into account the exogenous variables, which could be the determinants of efficiency. Bank management often rejects the results of the conventional efficiency studies as these may provide quite a biased or a single-perspective evaluation (Paradi et al. 2011). For example, when DEA recognizes one bank to be efficient, it might just be the case that the bank is in a more favorable environment than the other banks in the sample (Henrique et al. 2022), so efficiency scores can be impacted by factors that the bank administrators often have no control over. Therefore, finding out these external variables which might be affecting the efficiency scores, either positively or negatively, helps bank managers in taking better and well-informed decisions in that are in favor of the bank's financial health.

Considering all the different aspects of the DEA models as discussed above, it is preferred over the other efficiency techniques in this thesis. The external two-stage DEA models might be the most suited method for the research focus, which attempts to analyse the efficiency of EU banks, as well as find its determinants. But it is important to have a look at the previous empirical studies and their findings before confirming the exact model specifications for this thesis. Therefore, the next section will cover a review of papers that used DEA models to study bank efficiency, with a highlight on the two-stage models and EU bank efficiency.

1.2 Empirical studies

Řepková (2014) studied the efficiency 11 Czech commercial banks using an input-oriented DEA model for the period of 2003-2012. An intermediation approach was used for this analysis, based on which the input variables were labor and deposits, and the output variables were loans and net interest income. The results found that average efficiency calculated using the CCR model ranges from 70 % to 78 %, while with the BCC model it reached 84–89 % for the observed period. This difference in the average efficiency is always expected as the BCC model decomposes technical inefficiency into pure technical inefficiency and the inefficiency to scale, therefore not considering the portion of the inefficiency that is caused by small size of production units (Řepková, 2014). The author suggested that the reason for the Czech banking inefficiency could be due to the presence of excess client deposits on the balance sheet and inappropriate bank size choices.

One year later, the author conducted another study, Řepková (2015), extending the work from the previous paper, where an external second stage was added to the simple DEA model to find the efficiency determinants of Czech banks. The result shows that while bank size, credit risk, number of branches or concentration of the banking sector were not statistically significant. GDP had a significant positive influence on banking efficiency in both CCR and BCC models. Capitalization rate as measured by equity over total assets, interest rate, and profitability were found to be significant only in the CCR model, of which, the level of capitalization had a positive association with efficiency, while the latter two had influenced efficiency negatively.

Milenkovic (2021) also applied a two-stage DEA model for banks in the Western Balkan countries, with Tobit regression in the second stage to investigate the environmental factors that might have impacted the estimated efficiency scores. An output-oriented BCC model was applied on the first stage to measure the technical efficiency of banks. Consistent with the intermediation approach, deposits, labor costs, and bank capital were the inputs, while loans and investments were the outputs in the first stage. The author found that the overall score of efficiency for banks in the selected countries was above 85%. The environmental factors in the second stage were dummy variables of bank type (commercial or investment), bank size, and merger and accusations (M&A).

While bank type and M&A had significant negative influence on efficiency level, bank size on the other hand had significant positive association. The results also suggest that larger banks are more efficient due to having a higher deposit potential. Moreover, a higher efficiency for commercial banks was detected as expected by the author because they are more focused on generating placements from deposits, unlike investment banks which rely on non-deposit sources of financing (Milenkovic 2021).

Casu and Molyneux (2003) conducted a comparative study between the European countries for the period 1993-1997 to investigate if the productive efficiency of banks improved and converged toward the single market. Majority of the banks in the sample had an efficiency level of approximately 65%, and the bank efficiency in all countries, except Italy, showed significant improvements within the observed period. Moreover, using both bootstrap and Tobit regression in the second stage of the DEA analysis, the authors found strong evidence that country-specific aspects play a significant role in explaining the efficiency differences across European banking systems. Factors such as the use of different banking legislations and managerial strategies to address the new difficulties introduced by technological innovation, and increased competition within the EU banking industry can be some of the reasons behind this efficiency difference (Casu, et al. 2003). Besides this, a negative association was also found between banking efficiency and other external variables such as, average capital ratios and return on average equity.

Most recently, Anagnostopoulos et al. (2022) conducted a three-phase comparative efficiency analysis between American and European banking sector over the period 2000-2018. An output-oriented DEA model with an intermediation approach was used in the first stage. The findings of the first stage suggests that although the EU banks had significantly lower levels of efficiency compared to American banks, the growth rate for efficiency post financial crisis for were much higher for the EU banks (16%) than the American banks (6%). The study also adapted a difference-in-difference estimation (DID) model in the second stage to estimate the terminal effects due to policy or other economic environment changes and it is based on the comparison between a treatment group and control group (Anagnostopoulos et al. 2022). Finally, the third stage Tobit estimations suggested that an increase in bank size and level of equity led to a higher level of technical efficiency for banks from both regions, whilst a higher loans to asset ratio was negatively associated with the EU bank efficiency. And lastly, macroeconomic variables had a stronger impact on smaller banks. Overall, the lower level of bank efficiency for EU could be related to the larger size of EU banking sector as a percentage of GDP compared to the US, which means any regulatory mistakes will be more severe for the EU banks (Anagnostopoulos et al. 2022).

It is worth pointing out that the use of efficiency analysis is not limited to only bank efficiency estimations. Lin et al (2009) studied the operating efficiency of 117 branches of a Taiwanese bank using DEA. The input variables used in this study were number of employees, interest expense, and deposit, while the output variables were amount of loan, interest revenue, operating revenue and earnings. Overall technical efficiency scores under the CCR model shows that nearly half of the studied DMUs (branches) were inefficient. Pure technical efficiency score was 0.67, so 33% of the cost was wasted, which supposedly could be due to lower loan-to-deposit ratio causing an excessive input waste or the existence of a high number of local banks. Scale efficiency estimates also suggests that the resource waste ratio due to technical inefficiency (0.452) might be due to bad managerial choices or poor performance executions. The empirical results suggested that although business size does not impact the quality of efficiency directly, supervisory officers can use the orientation and size of the branch unit as a baseline to improve performance assessment of its employees (Lin et al. 2009). Overall, the study shed light on the importance of ranking DMUs based on efficiency levels to achieve the incentive effect and to evaluate management performance more objectively.

Barth and Staat (2005) also analysed the efficiency of 31 branches from a German bank and verified the impact of external variables such as branch area, public transport, competition and others on branch efficiency. Branch managers often use these factors as an excuse to convince the higher management that poor bank performance is influenced by the external factors that are outside of managerial control (Barth, et al. 2005). However, the result of a bootstrapped truncated regression in the second stage of DEA verified that environmental factors do not impact bank branch efficiency as none of the variables were statistically significant. Thus, using a two-stage approach in this study allowed for a more elaborate analysis to fulfill the aim of their research, one of which was whether external factors determine bank efficiency or not.

Based on the studies mentioned above, regression analysis seems to be the most common second stage method when using an external DEA model. But it is worth pointing out that the second stage in an external DEA model can have different methods and are not limited to only regression analysis. Wu et al. (2006) conducted a study analyzing the relative efficiency of 142 branches of a large Canadian bank. They applied DEA to measure bank efficiency and, at the same time, used these efficiency scores to train an Artificial neural network (ANN). Since DEA is quite sensitive to the existence of outliers, ANN was also considered as it can be used to find data envelopes supported by the whole dataset, instead of being based on outliers. It was concluded that although the empirical result in this study could be compared to traditional DEA models, DEA-NN approach was able to identify more efficient units by exploring good performance patterns (Wu et al. 2006)

2. DATA AND METHODOLOGY

2.1. Selection of variables

The aim of this thesis is to investigate the efficiency of European banks and determine the factors affecting the efficiency scores. Therefore, a two-stage external DEA method is adopted, where the efficiency scores are measured in the first stage through a DEA analysis, and the efficiency determinants are found in the second stage using a regression analysis. For the first stage DEA, the input and output variables need to be defined based on different approaches available in the literature on banking efficiency. Two of the most popular approaches to define the input-output relationship in financial institutions are the production and the intermediation approach.

The production approach, as proposed by Benston (1965), considers that bank's primary function is to provide services to its clients, and it is more appropriate for studies dealing with bank branch efficiency. Contrary to this, the intermediation approach proposed by Sealey, et al (1977) is widely used among literatures where the research focus is on the banks or financial institutions themselves (Henriques, et al. 2022). According to this approach, bank's main objective is to act as a financial intermediator between savers and borrowers (Hou et al. 2014). It assumes that banks collect deposits to transform them with labour and capital into loans (Chen et al. 2020). The choice of variables can largely impact the efficiency scores obtained (Henriques, et al. 2022), therefore the selection needs to be made to fit the research objective as closely as possible to ensure better accuracy of the results. Since the focus of this thesis is to analyse bank efficiency, an intermediation approach is more suitable. Consistent with the intermediation approach, the input (x) and output (y) variables will be selected for the first stage of the analysis.

Total deposits (DP) and Staff expenses (STAFF) are some of the most common variables used in the bank efficiency studies. While the latter have mostly been included as an input variable in the previous literatures, the treatment of deposits is rather controversial. Interestingly, the impact of deposits on the DEA model can vary a lot depending on whether it is treated as an input or output. Holod et al (2011) used deposits as an output following the production approach and found that when all other variables are kept constant, a higher level of deposit would lead to higher efficiency scores. Contrary to this, when deposits are used as an input, a lower deposit results in higher efficiency scores (Henriques, et al. 2022). Moreover, majority of the studies that adopted an intermediation approach, treated both the variables, total deposits, and staff expenses, as inputs (Řepková, 2014; Milenkovic, et al. 2022; Casu and Molyneux 2003). Therefore, these will be included as inputs for this thesis consistent of the intermediation approach. It is worth pointing out that a few studies have also used the total number of employees, instead of the total amount of staff expenses as an input (Anagnostopoulos et al. 2022; Zimoková et al. 2015). But since the number of full-time equivalent staff was not available for all the banks in the sample, this was not an option for this paper. Fixed asset (FE) is also used as an input variable in various bank efficiency studies that adopted an intermediation approach (Li 2020; Anagnostopoulos et al. 2022; Sufian et al. 2016). Finally, physical capital, as measured by other operating expenses (OOE) (Thoraneenitiyan et al. 2009; Zimoková, Bod'a 2015), is the final input selected for the first stage analysis.

Furthermore, the DEA output variables for this thesis are Total loans (TL), Other earning assets (OEA), and Off balance-sheet items (OFFBS). The amount of total loans can be considered as quite a traditional variable due to its inclusion in majority of the bank efficiency studies (Řepková 2014; Anagnostopoulos et al. 2022; Ferreira 2019; Chen et al. 2021). Additionally, contingent liabilities such as letters of credit, derivatives and other types of non-traditional activities are becoming increasingly important in European and global banking and excluding these items from the DEA model may lead to biased bank performance measurements (Casu et al. 2004; Thoraneenitiyan et al. 2009; Hassan et al. 2003). Consistent with this argument, the second output of this thesis is selected, which is the value of banks' off balance-sheet items. Moreover, Ferreira (2019) suggests that an increase in the amount off balance-sheet items is associated with risk transfer, liquidity enhancement, and regulation requirements, which would benefit the overall bank performance. Finally, the last output for the DEA model is other earning assets (OEA), which measures the performance of banks' portfolio management (Thoraneenitiyan et al. 2009; Ferreira, 2019, Chen et al, 2021; Blankson et al. 2022).

For the second stage, the external variables are selected to act as the independent variables against the technical efficiency scores found from the first stage. The chosen variables are consistent with the studies discussed in section 1.2. Bank size (SIZE), when measured by the natural logarithm of total assets, is a common and significant determinant of EU bank efficiency as suggested by the result of several studies. It is known to be responsible for controlling the economies of scale which provides an indication as to the impact of the bank's size on the efficiency score (Doan et al. 2018). Economies of scale refers to the phenomenon that in the long run, the larger a firm gets, the more it benefits from increasing returns to scale, and other cost-saving measures associated with the output production. Overall, bank size is expected to have a positive association with the level of efficiency as larger banks may benefit from its market powers, reduced risk, and economies of scale (Menicucci et al. 2015). However, it is worth pointing out that the findings of some studies suggest an opposite relationship. When Anagnostopoulos et al (2022) divided the European and American banks according to their size, they found that smaller banks in the sample had higher a higher efficiency technical efficiency scores than the medium ones, and the largest banks had the lowest scale efficiency.

The next external variable is the capital ratio, as calculated by the ratio of total equity over total assets (EQTA). It is a much-appreciated tool to assess a bank's capital strength and financial soundness (Menicucci et al. 2015). Moreover, capital ratio can be a great indicator of a bank's ability to withstand unforeseen dangerous situations like financial crisis or possible losses by measuring how well-capitalized a bank is (Anagnostopoulos et al. 2022). Many of the previous findings suggests that banks with higher capital ratios are more efficient due to facing lower costs associated with the financial distress (Casu and Molyneux 2003; Řepková 2015; Sufian et al. 2016; Zhang et al. 2014).

Along with the capital ratio, many efficiency studies also employ loans to total asset ratio (LOANTA) in their external regression models to capture the impact of liquidity risk, or act as a proxy for credit risk (Řepková 2015; Sufian et al. 2016; Anagnostopoulos, et al. 2022). The direction of its association with the bank efficiency is a bit difficult to predict due to the diverse findings from the previous studies. While a higher loan ratio should have a positive impact as loans are the main source of bank earnings, an extremely high volume of loan portfolio can also result in the drop of credit quality, resulting in poor performance (Menicucci et al. 2015). Contrary to this, too low of a value may also indicate that banks will have higher opportunity costs for not providing enough loans.

Similarly, loans to deposit ratio (LOANDP) can also be interpreted as another measure of liquidity risk. Some studies have employed both LOANTA and LOANDP in their regression models as the latter can reveal the additional information about if the bank requires non-deposit fundings to provide loans (Vodová 2014; Řepková 2015). When this ratio lower than 100%, it suggests that the all the bank loans are funded by the deposits. Deposits are considered as a bank liability and loans are the main source of bank earnings, the difference between the interest rate paid to the depositors and the interest rate charged for the loans provided is mainly the bank's revenue. Therefore, a higher LOANDP is expected to have a positive association with bank efficiency due to the increase in profits associated with higher lending (Lee et al. 2014). However, higher values are also associated with a greater risk because when banks do not have enough deposits, they need to rely heavily on non-deposit funding sources that are often quite sensitive to economic conditions, both in terms of availability and priciness (Disalvo et al. 2017).

Finally, the macroeconomic variable, GDP per capita (PCGDP), is the last variable for the regression model, consistent with the previous studies (Řepková 2015; Anagnostopoulos et al. 2022; Chen et al. 2021). In this thesis, it will be added as a control variable to check whether it contributes to any changes in the efficiency levels, and the overall regression models. As suggested by Chen et al (2021), PCGDP may have a positive and significant impact on bank efficiency as a higher per capita income associated with the bank's ability to attract more deposits and generate stronger cash flow.

2.2. Sample data collection and descriptive statistics

The bank level data for this paper has been gathered from the database BankFocus, and the macro specific data is collected from Worldbank. As the focus of this paper is on the EU banks, the final dataset consists of bank data from 9 different EU countries. Data is observed from 2010-2020. The most recent year, 2021 was not included as many of the banks did not have the needed data available yet for this year. The dataset needed to be cleaned before conducting the analysis. Therefore, all the bank years with missing or not available values were removed. Moreover, DEA is highly sensitive to extreme values, therefore, the outliers from the dataset were also removed using the 1.5*IQR test. IQR represents the interquartile range, and the test mainly considers values as outliers if they either fall below the value given by this equation, Q1-1.5*IQR, or above the value given by this equation, Q3+1.5*IQR, where Q1 and Q3 are the first and third quartile values, respectively.

After the removal of missing values and the outliers, the final dataset contains 3952 observations from 654 banks belonging to 9 EU countries during the 11-year period of 2010-2020. The initial goal was to also include countries such as Czech Republic, Belgium and Ireland. But after the whole dataset was cleaned, the number of remaining banks were too low to make any realistic estimates, hence they were removed as well. The number of banks selected from each of the countries are presented in Table 1, which shows that the sample contains the largest number of banks from Germany, followed by Austria and France. The final sample can be considered as a panel dataset as it contains both time and cross-sectional observations, from different companies over different periods of time, but the removal of several bank years during the data cleaning process resulted in an unbalanced panel dataset.

Table 1. Number of banks by country

Country	Number of banks.
Austria	64
Germany	366
Denmark	29
France	48
United Kingdom	29
Italy	42
Luxembourg	22
Poland	13
Sweden	41

Source: Author's own calculations

As discussed in 2.1, in total, four input and three output variables is selected for the DEA model, while six independent (external) variables, including a macro specific control variable, are chosen for the regression models. Table 2 summarizes all the measurements made using the collected data, as well as the abbreviations of the variables.

Table 2. Variables abbreviations and measurements

Variables	Abbreviations	Measures
First stage: DEA analysis		
Inputs (x):		
Fixed assets	FA	Fixed assets net of accumulated depreciation
Total deposit	DP	Total deposits from customers + Interbank deposits and amounts due to other financial institutions
Staff expenses	STAFF	Employee expenses + post-employment benefits
Other operating expenses	OOE	Expenses related to the bank's operations other than Staff and administrative expenses
Outputs (y):		
Total loan	LOAN	Net loans and advances to customers and banks
Other earning assets	OEA	On-balance sheet assets that generate non-interest earnings for the financial institution
Off balance-sheet items	OFFBS	Sum of all off-balance sheet exposures
Second stage: Regression		
Independent variables:		
Size of the bank	SIZE	$\ln(\text{total assets})$
Loans to asset ratio	LOANTA	total loans/total assets
Equity ratio	EQTA	total equity/total assets
Loans to deposit ratio	LOANDP	total loans/total deposits
GDP per capita (constant US$)	PCGDP	GDP/total population

Source: Compiled by author

To better understand the characteristics of the sample dataset, Table 3 provides the descriptive statistics of the variables. All the variables are in million euros, and most of the external variables are calculated as ratios. The mean value of the total deposit for the 654 banks analysed is €1,409 million, with a standard deviation (SD) score of 1,023.27. The high value of standard deviation indicates that European banks are quite dispersed in terms of total deposits from customers. Similarly, the mean and standard deviation values were quite high for most of the outputs, which was expected for the selected sample as it contains banks from 9 different countries with varying sizes. This is consistent with the findings of Anagnostopoulos et al (2022) for EU banks from 28 countries. Moreover, all three of the output variables, total loans, other earning assets and off balance-sheet items, had high SD, maximum (Max), and minimum (Min) values, suggesting that the mean value of these variables are quite far away from the true representation for majority of the banks in the sample.

Table 3. Descriptive statistics of all the variables

	Mean	Median	SD	Min	Max	N
DEA variables:						
Fixed assets	13.06	10.43	11.57	0	51.89	3952
Total deposits	1,409.45	1,241.04	1,023.27	62.67	5,182.07	3952
Staff expenses	18.21	15.73	12.95	0	64.71	3952
Other operating expenses	6.90	5.03	6.59	-0.08	28.77	3952
Total loans	1,171.61	988.86	883.28	32.42	4,501.09	3952
Other earning assets	504.63	408.33	413.96	0.03	1,929.89	3952
Off balance sheet items	136.47	99.24	125.61	0	590.55	3952
Environmental variables:						
Size	7.064	7.270	0.920	4.469	8.683	3952
Equity to asset	0.092	0.091	0.025	0.025	0.159	3952
Loan to asset	0.620	0.636	0.148	0.198	0.997	3952
Loan to deposit	0.845	0.841	0.003	0.282	3.915	3952
Per capita GDP	0.106	0.106	0.002	0.093	0.116	3952

Source: Author’s own calculations based on data from BankFocus

As for the external variables, the average value of loan to asset (LOANTA) is 0.620. This is close to the findings of Anagnostopoulos et al (2022) (0.553) for EU banks who suggests that LOANTA is expected to have a negative association with efficiency levels as the higher the ratio, the riskier a bank may be to higher defaults. Moreover, the loan to deposit ratio for all banks in this thesis averaged to 0.845. While a higher loan to deposit ratio is expected to increase bank profitability, the government in some countries put a limit to this ratio in order to reduce the risk associated with higher lending, as discussed in 2.1. China's banking law for example has a put of 75% at the total loan to deposit ratio (Guonan 2014).

Table 4. Correlation matrix of external variables

	SIZE	EQTA	LOANTA	LOANDP	PCGDP
SIZE	1				
EQTA	-0.19	1			
LOANTA	0.09	-0.06	1		
LOANDP	-0.04	-0.09	0.58	1	
PCGDP	-0.08	0.15	-0.09	-0.05	1

Source: Authors own calculation using Microsoft Excel

Lastly, the correlation matrix of the external variables to be used in the second stage is presented in Table 4. A Pearson correlation matrix can identify the bivariate relationship or the strength of association between two independent variables. It is highly important to check the level of correlation between the independent variables before conducting multiple regression analysis. Multicollinearity is an issue that arises when two, or more independent variables in the sample dataset are highly correlated. Such situations can hamper the quality of the analysis by providing biased or inaccurate results, therefore those independent variables need to be removed from the model. However, multicollinearity does not seem to be an issue in this model as the strongest correlation coefficient is found between LOANTA and LOANDP ratios (0.58). Moreover, a variance influence test (VIF) has also been conducted to confirm the non-existence of multicollinearity. VIF is considered as a better approach than the simple correlation matrix because it can not only identify the correlation of one independent variable with a group of other variables, but also can also indicate for which coefficients the collinearity is a problem (Hernandez, et al. 2007). The VIF for the *j*th independent variable is given by,

$$VIF_j = \frac{1}{(1-R_j^2)} \tag{2}$$

Where, R^2_j is the R^2 from the regression of the *j*th explanatory variable on the remaining explanatory variables (Hernandez, et al. 2007).

Table 5. VIF test

Variables	R_j^2	VIF_j
SIZE	0.061	1.065
EQTA	0.064	1.069
LOANTA	0.355	**1.551**
LOANDP	0.356	**1.552**
PCGDP	0.050	1.053

Source: Authors own calculations

The VIF test (Table 5) suggests that the highest values are 1.551 and 1.552, which are for loans to asset and loans to deposit ratios, respectively. These values are pretty low, hence it confirms that there is no need to remove any of these variables. Therefore, it can be concluded that multicollinearity is not going to be an issue for the regression analysis in this thesis.

2.3. Research methodology

Consistent with the aim of this thesis, a quantitative method needs to be applied which can investigate both the level of EU bank efficiency and its main determinants. Therefore, an external two-stage DEA analysis is the most suitable method to fulfil this purpose. This approach has mostly been adopted in studies that extended their research goal from only focusing on the efficiency scores to also find the various factors impacting those scores by the addition of a second stage (Henriques et al. 2022).

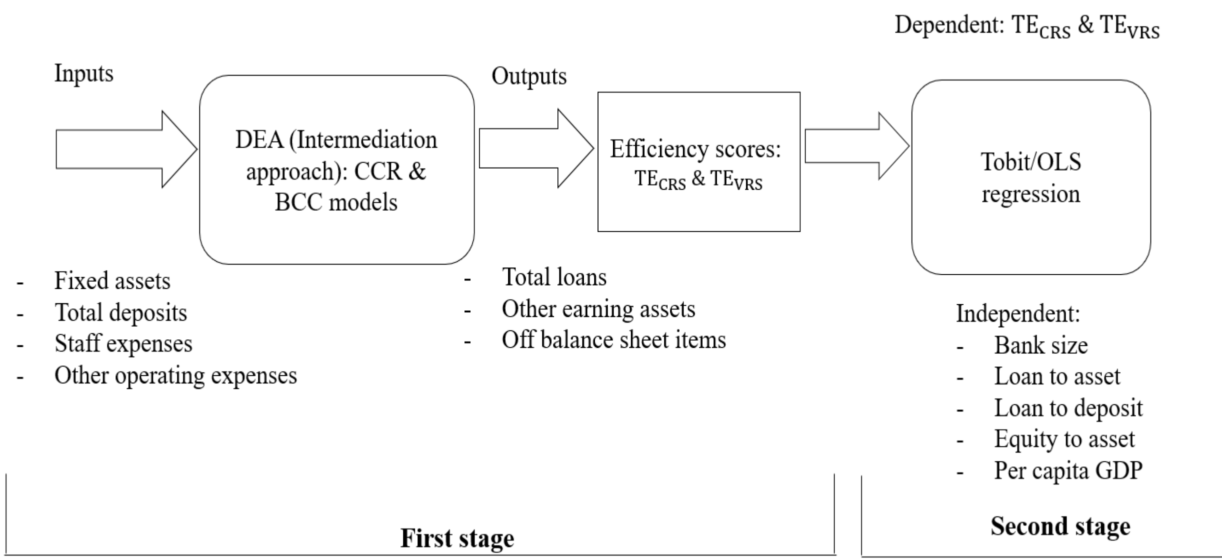

Figure 2. Diagram of the two-stage DEA model construction

Source: Author's own illustrations

The construction of the two-stage model is summarized in Figure 2. In the first stage, a traditional DEA model is used to measure the technical efficiency (TE) scores, which indicates the ability of the DMUs (bank's) to attain the highest level of output given the set of inputs. As already discussed in 2.1, the input and output variables have been selected consistent with the intermediation approach (Figure 2). Both BCC and CCR model will be used to calculate the efficiency scores under the TE_{CRS} and TE_{VRS} scores, respectively. These two efficiency scores generated from the first stage become the dependent variables for the second stage regression analysis against the selected external or independent variables.

An input-oriented DEA model is applied for this thesis, which assumes that banks tend to minimize costs (inputs) to become more efficient. Schaffnit et al. (1997, pp. 278) suggests that since banks do not have control over their outputs, input orientation should be more appropriate than output orientations. Moreover, many studies tend to select input-orientated measures because the input quantities appear to be the primary decision variables (Casu et al. 2003). The main difference between the orientations arises in the measurements of the technical inefficiencies. In the input-orientated DEA models, technical inefficiency is calculated as a proportional reduction in input usage, while in the output-oriented model, it is measured as a proportional increase in output production Casu et al. 2003). Therefore, both the input and output-oriented models will identify the same set of efficient DMUs, but the results will only vary for the efficiency measurements associated with the inefficient DMUs. However, the choice between which one of these models provides the best efficiency measurements remained inconclusive in the theoretical literatures to this date (Casu et al. 2003).

Since this paper will adopt an input-orientated model, the linear programming problem to solve for the DEA results under CRS assumptions is given by Coelli (1996). With the assumptions that the considered N banks (or DMUs) produce M outputs by using K inputs, so, X being the $(K \times N)$ input matrix, while Y being the $(K \times M)$ output matrix contain data of all the N DMUs (Ferreira 2019), the equation by Coelli (1996) can be presented as follows:

$$\begin{aligned} &\mathrm{Min}_{\theta,\lambda}\ \theta, \\ &\text{Subject to: } -y_i + Y\lambda \geq 0, \\ &\qquad \theta x_i - X\lambda \geq 0, \\ &\qquad \lambda \geq 0 \end{aligned} \tag{3}$$

Where,

X = K*N matrix of inputs

Y = M*N matrix of outputs

i = DMU or bank

x_i = DMUs' input and

y_i = DMUs’ output

θ = input distance measure

λ = N $\times$1 vector of constants

Solving (3) gives the efficiency score, θ, for each of the DMUs. In all cases, $\theta \leq 1$, and a θ=1 represents that a DMU is in the efficient frontier or is perfectly technical efficient. When DMUs are not in the frontier, (1- θ) represent the technical inefficiencies, or the distance to the frontier (Ferreira 2019).

The technical efficiency scores calculated under CCR model (TE_{CRS}) using (3) can be seen as a product of pure technical efficiency (TE_{VRS}) and scale efficiency, where the first may capture the managerial performance, while the latter indicates how closely a DMU operates to its optimal of production. The ratio of TE_{CRS} over TE_{VRS} captures the scale effects, as given by equation (1) in section 1.1. The BCC model calculates the pure technical efficiency with variable return to scale assumptions, TE_{VRS}. This requires an additional constraint $N1'\lambda = 1$ to be added to equation (3) above. The constraint for the BCC model, still following (Coelli 1996) is given as:

$$\begin{aligned}
&\mathrm{Min}_{\theta,\lambda}\ \theta, \\
&\text{Subject to: } -y_i + Y\lambda \geq 0, \\
&\qquad \theta x_i - X\lambda \geq 0, \\
&\qquad N1'\lambda = 1, \\
&\qquad \lambda \geq 0 \qquad\qquad (4)
\end{aligned}$$

Where,

X = K*N matrix of inputs

Y = M*N matrix of outputs

i = DMU or bank

x_i = DMUs' input and

y_i = DMUs' output

θ = input distance measure

λ = N $\times$1 vector of constants

N1 = N $\times$1 vector of ones

To determine the factors affecting bank efficiency scores in the second stage of the analysis, the preferred model will be formulated using the Tobit regression. The Tobit model is known as a censored regression which can censor for a specific value such as on the left (minimum) or right side (maximum) of the regression. When operating as part of a DEA process, where the maximum score is 1 and minimum score is 0, implantation of the Tobit model can provide better estimates of the results than other traditional regression models such as Ordinary least squares (OLS) (Anagnostopoulos et al. 2022). However, several studies including Řepková (2015) have implemented an OLS model to find the bank efficiency determinants. For the second stage of the analysis, Tobit regression model will be adopted, but OLS results will be provided, mainly for comparison and robustness check purposes.

Following Greene (2003, pp 764), the common specification for the Tobit model can be written as follows:

$y_{\mathrm{i}}^{*} = \mathrm{x}_{\mathrm{i}}'\beta + \varepsilon_{i}$,

y_{i} = 0 if $y_{\mathrm{i}}^{*} \leq 0$

$y_{i} = y_{\mathrm{i}}^{*}\ if\ y_{\mathrm{i}}^{*} \geq 0$ (5)

Where,

y = dependent variable

y* = latent dependent variable of the technical efficiency result for positive values and censored otherwise

x′ = vector of explanatory variables

β = vector of estimable coefficients,

ε = normally and independently distributed error term

i = DMU or bank index

Finally, equation for the pooled OLS can be written as given by Greene (2003, pp 11):

$$Y_{i} = \beta_{0} + \beta_{1}X_{1} + \beta_{2}X_{2} + \beta_{n}X_{n} + \varepsilon \quad (6)$$

Where,

Y_i = dependent variable

β = estimated coefficient

$\beta_{(1,2,\ldots n)}$ = coefficient corresponding X_n variables

$X_{(1,2,\ldots n)}$ = independent variables

ε = Error term

3. FINDINGS AND DISCUSSION

3.1. First stage: DEA analysis results

The DEA analysis has been conducted in the R studio, using the package ‘‘Benchmarking’’. The summary of the efficiency scores calculated under both CCR and BCC models are presented in Table 6. The average pure technical efficiency for the whole sample under the BCC model (TE_{VRS}) is 45.9%, where 102 out of 3952 observations are fully efficient with the score of 1. Moreover, majority of the observations (approximately 60%) fall within the efficiency range of 0.30-0.50 (Appendix 1), and the third quartile value (Q3) indicates that only 25% of the observations have the efficiency score above 0.527. On the other hand, the average efficiency score with under the CCR model (TE_{CRS}) is 0.388, with majority of the observations having the efficiency scores lower than 0.420, as indicated by the Q3 value. Under the CRS assumption, only 25 observations were technically efficient (Appendix 1). Overall, the results suggest that, on average, technical efficiency scores are rather low for the selected EU banks during 2010-2020 using both models.

Table 6. Efficiency scores

	TE_{CRS}	TE_{VRS}	Scale
Mean	0.388	0.459	0.876
Median	0.355	0.406	0.931
Standard Deviation	0.127	0.174	0.143
Minimum	0.076	0.188	0.105
Maximum	1.000	1.000	1.000
1^{st} quartile (Q1)	0.308	0.337	0.826
3^{rd} quartile (Q3)	0.420	0.527	0.977
Observations	3952	3952	3952

Source: Authors own calculations using based on R results

The average TE score calculated using the BCC model is higher than the CCR model. This was expected, and it also aligns with the previous findings since TE_{VRS} is a measure of pure technical efficiency, which unlike the $\mathrm{TE_{CRS}}$, does not consider the scale inefficiency effects. As already discussed in 1.1, a difference between the $\mathrm{TE_{CRS}}$ and $\mathrm{TE_{VRS}}$ scores of a DMU indicate the existences of scale inefficiencies (Sufian 2007), which seems to be the case for most of the observations in this thesis. Scale efficiency score, when measured as the ratio of $\mathrm{TE_{CRS}/TE_{VRS}}$, ranges between 0 and 1, and the closer a bank's score is to 1, the more scale efficient it is (Majumdar et al. 1996). Average scale efficiency for all the banks among the period under observation is 0.876, which is quite high and indicates that banks in the selected EU countries are operating not so far away from their most productive scale size.

As it can be seen in Figure 3, the efficiency scores do not vary much between 2010-2020. The only biggest jumps can be observed during 2010-2012. Both $\mathrm{TE_{VRS}}$ and $\mathrm{TE_{CRS}}$ rises by 3% in 2011 compared to the previous year. As Anagnostopoulos, et al (2022) explains, this increase could be due to the ability to the Europeans banks’ ability to recover post the global financial crisis. But interestingly, the efficiency scores drop back in 2012 and return to almost the same level as it was before the rise (Figure 3). Followed by this drop, the score range was quite small during 2012-2020, with minimal changes that did not exceed over 2%. The scale efficiency was rather stable during the entire period, with the lowest score being 0.86 in 2010 and the highest being 0.89 in 2012. These trends can be compared with the findings of with Ferreira (2019) who analysed the EU bank efficiency between 2011-2017. Evidence of a steady decrease in both the TE measures was found between 2011 and 2012, which the banks did not recover fully during the rest of the interval.

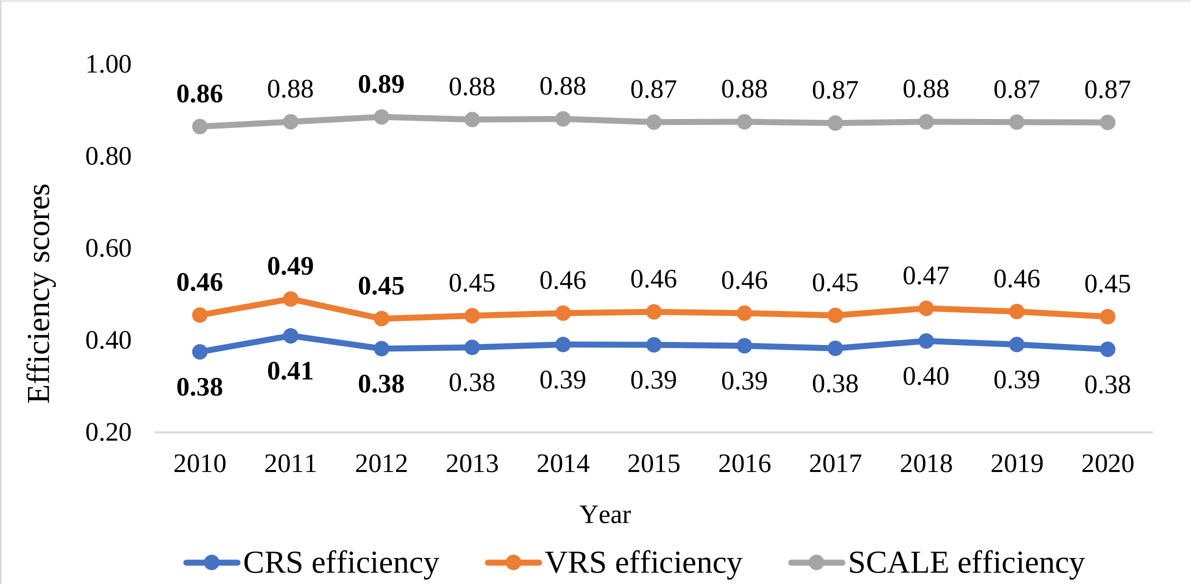

Figure 3. Average efficiency score from 2010-2020

Source: Compiled by author using Microsoft Excel

The technical efficiency scores calculated for the whole sample are rather low compared to previous findings. Anagnostopoulos et al (2022) found that for the period 2000-2018, the average TE_{VRS} score for 52 EU banks from 28 countries was 0.705, while the TE_{CRS} score was 0.532. Řepková (2015) found even higher values for Czech banks during 2001-2012 under both the CCR (0.754) and BCC (0.922) models. Finally, the findings of Ferreira (2019) using an input-oriented DEA model suggests that the mean efficiency scores for 485 EU banks during 2011-2017 were 0.625 and 0.685 under CCR and BCC models, respectively. One of the reasons behind this difference in average scores for the overall sample could be explained by the variation in bank efficiency levels between the selected countries. As suggested by Casu and Molyneux (2003), country-specific characteristics play a significant role in the explaining of bank efficiency levels among EU countries. Therefore, varying degrees of bank efficiency in different countries might have impacted the average scores of the overall sample.

Table 7 presents the average efficiency scores separated by the countries, where it can be clearly seen that the bank efficiency levels are indeed highly dispersed among the selected EU countries in this thesis. Banks from some countries have a significantly lower TE scores than others, which indeed impacted the overall sample scores as well. Based on the average TE_{VRS} scores for the period of 2010-2020, only Danish (0.621), Italian (0.605), and Luxembourgish (0.598) banks have relatively higher efficiency scores, while Polish banks have the highest score of 0.707. German and Swedish banks seem to have the lowest TE_{VRS} and TE_{CRS} scores, respectively. But it is also worth pointing out that the number of German banks analysed in this paper are much higher than the other countries. Hence, it could also be the case that the German banks in the sample varied a lot in size, and bank size being one of the potential determinants of efficiency, might have resulted in the lower average scores for Germany.

Some of these results from Table 7 are closely supported by the findings of Ferreira (2019) during 2011-2017, where the pure technical efficiency scores (TE_{VRS}) for banks from Denmark, Italy and Poland averaged to 0.642, 0.696, and 0.728, respectively. However, the author found much higher values for the rest of the countries. It might be because Ferreira (2019) assumes that banks use equity, interest, and non-interest expenses as inputs to produce loans, other earning assets and non-earning assets as outputs. Unlike this thesis, the author considered equity as an input, and did not include the total deposits as either an input or output in the DEA model. Although the author adopted quite a similar method to estimate efficiency, the overall input-output combination is different than this thesis, which implies that the choice of variables can highly impact the level of efficiency, and it should be taken into consideration when analysing DEA results.

Similarly, when analysing the bank efficiency of 5 EU countries using a slightly different set of DEA variables than this thesis, Casu et al. (2003) also found higher efficiency scores for the whole sample under both BCC and CCR models. The period under observation in their paper (1993-1997) is almost over a decade apart from the one considered this thesis, which could be another reason behind the difference. But overall, the findings of both Casu et al. (2003), and Ferreira (2019) suggest high variation in bank efficiency levels between the EU countries, which is also the case in this thesis. As explained by Casu et al. (2003), the rising popularity of information technology and other financial innovations to digitalise the EU banking sector, along with a greater competition within the industry, can all lead to variations in the banking regulations and managerial choices in different countries. Therefore, these country-specific aspects of the banking technology could also be one of the reasons behind the efficiency differences among the countries in this thesis.

Table 7: Average efficiency scores separated by countries

Country	No. of banks	TE_{CRS}	TE_{VRS}	Scale
Austria	64	0.457	0.575	0.825
Denmark	29	0.504	0.621	0.844
France	48	0.441	0.501	0.900
Germany	366	0.347	0.404	0.889
Italy	42	0.524	0.605	0.888
Luxembourg	22	0.598	0.660	0.900
United Kingdom	29	0.520	0.583	0.902
Poland	13	0.573	0.707	0.833
Sweden	41	0.372	0.542	0.745

Source: Authors own calculations, based on RStudio results

In order to check if the estimated results of efficiency for the whole sample was impacted by any possible biases to the differing bank sizes, the sample banks have been divided into three categories based on their values of total assets (TA) in million euros. Based on the first and third quartile values of total assets, the smallest 25% of the bank observations (TA ≤ €680) represent Group 1, while the largest 25% of the bank observations (TA ≥ €2279) fall into Group 3 category. Finally, Group 2 represents all the observations after removing the smallest and largest ones from the sample (€680<TA<€2279), therefore, might provide a better or unbiased efficiency estimates. Each of the defined groups have been run separately in R Studio to get the DEA efficiency scores, the results of which are presented in Table 8.

Table 8. Efficiency estimates by bank size

	$TE_{(CRS)}$	$TE_{(VRS)}$	Scale	Observations
Group 1 (TA ≤€680)	0.576	0.685	0.841	988
Group 2 (€680<TA<€2279)	0.527	0.579	0.910	1976
Group 3 (TA ≥ €2279)	0.417	0.490	0.851	988

Source: Authors own calculations based on the DEA results from RStudio

As the results suggest, the technical efficiency scores without the presence of the smallest and largest bank observations (Group 2) are much higher than what was for the full sample. Group 2 is also the most scale efficient, indicating that banks of this size range are operating closest to their optimal production scale than banks in the other groups. This result is supported by Anagnostopoulos, et al (2022), who also found that when European banks were categorised by size, the medium sized banks had the highest scale efficiency score (0.905). Interestingly, the smallest banks (Group 1) scored the highest in terms of the average TE scores, while the largest ones (Group 3) were the least technically efficient, although both groups had similar scale efficiency scores. A broader perspective on how bank size impacted the overall sample of this thesis can be provided through the second stage regression results in the next section. But the results in Table 8, at the least, provide strong evidence about the sensitive nature of the DEA estimates, even after the removal of the extreme value or outliers from the sample.

3.2. Second stage: Regression results

After the technical efficiency scores are calculated in the first stage under both the CCR and BCC models, these will become the two dependent variables against the external variables in the second stage of the analysis, where both Tobit and OLS regression will be performed. The Tobit regression was conducted in RStudio using the package ‘‘censReg’’. The results of the Tobit and OLS regression can be seen below, in Table 9 and 10, respectively.

Table 9. Tobit regression results

Model	(9.1)	(9.2)	(9.3)	(9.4)
Dependent variable	TE_{CRS}	TE_{VRS}	TE_{CRS}	TE_{VRS}
Intercept	0.545***	0.886***	0.535***	0.977***
	(0.016)	(0.023)	(0.079)	(0.118)
SIZE	-0.012***	-0.055***	-0.012***	-0.055***
	(0.002)	(0.002)	(0.002)	(0.002)
LOANTA	-0.707***	-0.774***	-0.707***	-0.776***
	(0.013)	(0.019)	(0.013)	(0.019)
EQTA	0.058	0.443***	0.057	0.452***
	(0.013)	(0.089)	(0.060)	(0.090)
LOANDP	0.426***	0.473***	0.426***	0.474***
	(0.009)	(0.0144)	(0.009)	(0.014)
PCGDP			0.093	-0.854
			(0.726)	(1.082)
Observations	3952	3952	3952	3952

Source: Compiled by author based on Tobit results from R Studio

Note: 1) This table shows the results of Tobit regression models for both dependent variables

2) Standard errors are in parenthesis

3) ***p<0.01, **p<0.05, *p<0.10

Table 10: Pooled OLS regression results

Model	(10.1)	(10.2)	(10.3)	(10.4)
Dependent variable	TE_{CRS}	TE_{VRS}	TE_{CRS}	TE_{VRS}
Intercept	0.563***	0.909***	0.566***	0.988***
	(0.015)	(0.022)	(0.079)	(0.115)
SIZE	-0.013***	-0.056***	-0.013***	-0.056***
	(0.002)	(0.002)	(0.002)	(0.002)
LOANTA	-0.682***	-0.725***	-0.682***	-0.726***
	(0.012)	(0.018)	(0.012)	(0.018)
EQTA	0.052	0.414***	0.052	0.422***
	(0.060)	(0.087)	(0.060)	(0.088)
LOANDP	0.396***	0.419***	0.396***	0.419***
	(0.009)	(0.012)	(0.009)	(0.012)
PCGDP			-0.035	-0.741
			(0.726)	(1.055)
Observations	3952	3952	3952	3952
R^2	0.484	0.416	0.484	0.416

Source: Compiled by author using Microsoft Excel

Note: 1) This table shows the results of OLS regression models for both dependent variables

2) Standard errors are in parenthesis

3) ***p<0.01, **p<0.05, *p<0.10

The Tobit regression results suggest that all the external variables except for the equity ratio, are significant determinants of TE_{CRS} (Model 9.1). But when the external variables are regressed against TE_{VRS}, all of them were highly significant with a p-value lower than 0.01 (Model 9.2). The OLS results in Table 10 suggest that these associations do not vary much from Tobit estimations, so both the regression methods generate quite similar results. With the OLS models, R^2 value, also known as the coefficient of determinant, determines the proportion of variance in the dependent variable that can be explained by the independent variable. Simply put, R^2 value shows how well the data fit the regression model. As it can be seen in Table 10, the R^2 value for the models 10.1 and 10.2 are 0.419 and 0.485, respectively, which is not very hight and indicate that the selected external variables do not explain all the reasons behind the bank efficiency scores obtained in this thesis. Finally, Per capita GDP has been included as a control variable. Models 9.3 and 9.4 are the Tobit estimates, while 10.3 and 10.4 are the OLS results after the inclusion of this variable. But it was neither a significant variable in both OLS and Tobit models, nor did its inclusion majorly impacted the overall models.

When discussing the impact of each of the external variables on the efficiency scores, mainly Tobit results (Model 9.1 and 9.2) will be used for comparison purposes with previous studies, as this was the preferred method for this paper. But OLS results will also be brought to attention at times, especially when discussing the robustness tests.

The regression coefficient measures the strength of the relationship between the independent and dependent variables. Size is a negative and significant determinant of both $\mathrm{TE}_{(\mathrm{CRS})}$ and $\mathrm{TE}_{(\mathrm{VRS})}$, with coefficient values of -0.012 and -0.055, respectively. This was expected for the sample banks in this thesis as discussed in 3.1, where the average efficiency levels for the largest banks were the lowest. This result somewhat contrasts against some of the previous empirical findings. Studies such as Sufian (2016) and Milenkovic (2022) found a positive and significant association between size and bank efficiency. A possible explanation to support the findings is provided by Menicucci et al. (2015), which is that as banks keep getting larger, there comes a point beyond which they start facing higher costs per unit of production. This happens due to diseconomies of scale, where the economies of scale no longer functions. Lastly, the management of extremely large firms may also face higher expenses related to overheads of bureaucratic processes and agency costs, which overall, can prevent the bank to perform efficiently (Stiroh, Rumble 2006; Athanasoglou et al. 2008).

Loans to asset ratio, which is an indicator of liquidity risk, has a significant and highly negative association with the efficiency levels in this thesis. Several studies have found a positive association of LOANTA ratio with efficiency scores (Řepková 2015; Lee et al. 2015). Since loans are the main source of income for banks, a high ratio should benefit the bank performance. When banks rapidly increase their loans portfolio, it can lead often to an overall increased costs if the funding provisions needed to pay for the growth were higher than the profits earned (Menicucci et al. 2015), leading to a negative association with efficiency. Anagnostopoulos et al (2022) also found a significant and negative association between LOANTA ratio and EU bank efficiency (-0.321). Overall, it can be concluded that European banks could experience a lower efficiency estimate with an increased loans to asset ratio.

Equity to total asset ratio, which measures the strength of capital structure to bear losses, is only a significant variable for the VRS technical efficiency estimates. The results suggest a positive association with a coefficient value of 0.443 (Model 9.2), which aligns with the previous findings on EU bank efficiency (Casu, et al. 2003; Řepková, 2015; Anagnostopoulos, et al. 2022). This was expected as banks with higher capital ratios can survive without external fundings in times of financial crises. As suggested by Abreu et al. (2002), well capitalized banks from some of the European countries face lower costs relating to predicted bankruptcy and higher interest margins on profitable asset, therefore increasing the level of bank efficiency. On the other hand, banks with low capital ratio might not be able to overcome unstable macroeconomic conditions due to possible losses and not being able to afford external funding, resulting in a risk of insolvency (Menicucci et al. 2015). Overall, the result from this thesis confirms that well-capitalized EU banks are more efficient.

Lastly, loan to deposit ratio has a positive and significant association with the bank efficiency in this thesis. This is supported by the findings of Řepková (2015), Vodová (2015) and Lee et al. (2014). The results therefore confirm that a higher loan to deposit ratio can be advantageous for banks to be profitable. Loans generate revenue for banks in the form of interests paid by customers, hence the higher the lending, the more efficient a bank is. But banks should also be very careful with excessive lending and running out of deposit fundings. Extremely high loans to deposit ratio will require banks to borrow external or non-deposit fundings, especially in the times of financial downturns. These external fundings might not be easily available during period of crises, or might be overly expensive, all of which will eventually result in increased costs for banks, therefore, may eventually lower the efficiency levels (DiSalvo et al. 2017).

Lastly, three robustness checks were also conducted by removing the most important external variables as suggested by the findings, SIZE, LOANTA and EQTA. Only OLS regression was conducted for these tests to see if the removal of any of these variables impact the overall models. The result of these checks indicates that except for LOANTA, the removal of the other variables does not cause drastic changes in the overall models. A more detailed view of some of these results can be found in appendices 2 to 4.

CONCLUSION

The aim of this thesis was to investigate the efficiency of European banks, as well as the main determinants of the varying efficiency levels. In total, 654 banks from 9 different EU countries were selected. An external two stage DEA analysis was used to answer the two research questions, which were:

- How efficient is the banking sector in the selected European countries?
- What are the determinants that influence the different levels of bank efficiency?

The thesis fulfilled its aim by answering both the questions. To answer the first question, the result of the first stage of the analysis or the DEA estimates suggest that overall, the technical efficiency scores for the selected EU banks were quite low. The mean technical efficiency score using the CCR model was 38.8%, while with the BCC model it was a bit higher, 45.9%. But a high value of the average scale efficiency score indicated that the EU banks are more or less operating closer to their optimal production or scale size. Moreover, when separate DEA tests were conducted after dividing the banks into three groups based on their size, the overall efficiency scores were improved. The group containing the smallest banks were the most technical efficient, while the largest ones were the least efficient. This finding provided strong evidence that smaller European banks are more efficient. The result of the Tobit estimations in the second stage further confirmed that bank size had a significant and association with the level of bank efficiency.

Moreover, the findings from the regression analyses also suggest that higher loans to asset ratio resulted in lower bank efficiency for the selected EU banks in thesis. However, both the capital ratio and loan to deposit ratio had a positive and significant relationship with EU banks. It was also found out that the macroeconomic variable, per capita GDP did not have any significant impact on the efficiency scores.

Overall, the low level of bank efficiency for the overall sample in this thesis might be due to the following reasons. Firstly, the efficiency levels vary quite a lot between the countries. For example, Polish and Danish banks are relatively more efficient than German banks. This is supported by the findings of Casu et al. 2003, and the reason behind the difference is probably relating to the varying banking regulations among the countries that are brought forth to implement new banking technologies and financial innovations. The clear existence of bank inefficiency in the sample banks could also be due to inefficient managerial performances. Larger banks are most likely suffering from diseconomies of scale, which is an indication towards poor managerial strategies implemented on the selected EU banks. Therefore, bank administrators need to come up with improved strategies to reduce the cost per unit productions, which can help the larger banks to at least be more scale efficient. Finally, a negative impact of loan to asset ratio on the bank efficiency scores provides strong evidence of liquidity risk. Hence, EU banks need to be very careful when lending too much loans to customers in hopes of becoming more profitable.

To conclude, the research conducted for this thesis have a few limitations. Firstly, there is a large difference in the number of banks from the selected countries, which might have resulted in biased efficiency estimates. Some countries like Poland had very few banks in the final sample, while Germany had in total 366 banks. Therefore, a proper comparison between the countries was also not possible for the sample banks. Future research could extend the findings of this thesis by selecting a better dataset consisting of similar number of banks from each the countries. Moreover, bank types and industry specific variables were not considered as the potential determinants of EU bank efficiency in this thesis, mainly due to the lack of data availability. But previous findings such as Milenkovic (2022) suggests that bank types can significantly impact the efficiency levels. Hence, the future research papers can also enhance their findings by including these variables into the model.

LIST OF REFERENCES

Abreu, M. and Mendes, V. (2002), "Commercial bank interest margins and profitability: evidence from E.U. Countries", University of Porto Working paper Series, No. 122, available at: www.iefs.org.uk/Papers/Abreu.pdf

Alber, Nader and Elmofty, Mayssa and Kishk, Israa and Sami, Reda, Banking Efficiency: Concepts, Drivers, Measures, Literature and Conceptual Model (2019). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3310982

Anagnostopoulos, Y., Husa, K. A., & Noikokyris, E. (2022). A three-phase comparative efficiency analysis of US and EU Banks. *International Review of Economics & Finance*, *81*, 113–127. https://doi.org/10.1016/j.iref.2022.04.009

Aggelopoulos, E., & Georgopoulos, A. (2017). Bank branch efficiency under environmental change: A bootstrap DEA on monthly profit and loss accounting statements of Greek retail branches. *European Journal of Operational Research*, *261*(3), 1170–1188. https://doi.org/10.1016/j.ejor.2017.03.009

Athanasoglou, P. P., Brissimis, S. N., & Delis, M. D. (2008). Bank-specific, industry-specific and macroeconomic determinants of Bank Profitability. *Journal of International Financial Markets, Institutions and Money*, *18*(2), 121–136. https://doi.org/10.1016/j.intfin.2006.07.001

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Banker, R. D., Chang, H.-H., & Majumdar, S. K. (1996). Profitability, productivity and price recovery patterns in the U.S. Telecommunications Industry. *Review of Industrial Organization*, *11*(1). https://doi.org/10.1007/bf00163594

Berger, A. N., & Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, *98*(2), 175–212. https://doi.org/10.1016/s0377-2217(96)00342-6

Boďa, M., & Zimková, E. (2015). Efficiency in the Slovak banking industry: A comparison of three approaches. *Prague Economic Papers*, *24*(4), 434–451. https://doi.org/10.18267/j.pep.546

Benston, G. J. (1965). Branch banking and economies of scale. *The Journal of Finance*, *20*(2), 312–331. https://doi.org/10.1111/j.1540-6261.1965.tb00212.x

Blankson, N., Anarfo, E. B., Amewu, G., & Doabil, L. (2022). Examining the determinants of bank efficiency in transition: Empirical evidence from Ghana. *Heliyon*, *8*(8). https://doi.org/10.1016/j.heliyon.2022.e10156

Barth, W., & Staat, M. (2005). Environmental variables and relative efficiency of bank branches: A DEA-Bootstrap Approach. *International Journal of Business Performance Management*, *7*(3), 228. https://doi.org/10.1504/ijbpm.2005.006735

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of Decision Making Units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Casu, B., & Molyneux, P. (2003). A comparative study of efficiency in European banking. *Applied Economics*, *35*(17), 1865–1876. https://doi.org/10.1080/0003684032000158109

Chen, X., & Lu, C.-C. (2021). The impact of the macroeconomic factors in the bank efficiency: Evidence from the Chinese City Banks. *The North American Journal of Economics and Finance*, *55*, 101294. https://doi.org/10.1016/j.najef.2020.101294

Coelli, T.J. (1996) A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. CEPA Working Paper 96/08, University of New England, Armidale.

Esteve, M., Aparicio, J., Rodriguez-Sala, J. J., & Zhu, J. (2022). Random forests and the measurement of super-efficiency in the context of free disposal hull. *European Journal of Operational Research*, *304*(2), 729–744. https://doi.org/10.1016/j.ejor.2022.04.024

DiSalvo, J., & Johnston, R. (2017). *Banking trends: The rise in loan-to-deposit ratios: Is 80 the new 60?* Federal Reserve Bank of Philadelphia. Retrieved December 14, 2022, from https://www.philadelphiafed.org/the-economy/banking-and-financial-markets/banking-trends-the-rise-in-loan-to-deposit-ratios-is-80-the-new-60

Doan, A.-T., Lin, K.-L., & Doong, S.-C. (2018). What Drives Bank Efficiency? the interaction of bank income diversification and ownership. *International Review of Economics & Finance*, *55*, 203–219. https://doi.org/10.1016/j.iref.2017.07.019

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253. https://doi.org/10.2307/2343100

Førsund, Finn R. (2015). Productivity interpretations of the Farrell efficiency measures and the Malmquist index and its decomposition, Memorandum, No. 14/2015, University of Oslo, Department of Economics, Oslo. Retrieved from http://hdl.handle.net/10419/147432

Ferreira, C. (2020). Evaluating European bank efficiency using data envelopment analysis: Evidence in the aftermath of the recent financial crisis. *International Advances in Economic Research*, *26*(4), 391–405. https://doi.org/10.1007/s11294-020-09807-y

Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). *Biostatistics: A guide to design, analysis, and Discovery*. Elsevier Academic Press. Linear Regression (pp 349-386). https://doi.org/10.1016/C2009-0-03861-6

Greene, W.H. Econometric Analysis, 5th ed.; Prentice Hall: Hoboken, NJ, USA, 2003.

González, M. M., & Trujillo, L. (2009). Efficiency Measurement in the Port Industry: A Survey of the Empirical Evidence. *Journal of Transport Economics and Policy*, *43*(2), 157–192. http://www.jstor.org/stable/20466776

Guonan Ma. (2014) Tweaking China's loan-deposit ratio rule. Bruegel.

Henriques, I. C., Sobreiro, V. A., Kimura, H., & Mariano, E. B. (2020). Two-stage DEA in banks: Terminological controversies and future directions. *Expert Systems with Applications*, *161*, 113632. https://doi.org/10.1016/j.eswa.2020.113632

Hou, X., Wang, Q., & Zhang, Q. (2014). Market structure, risk taking, and the efficiency of Chinese commercial banks. *Emerging Markets Review*, *20*, 75–88. https://doi.org/10.1016/j.ememar.2014.06.001

Holod, D., & Lewis, H. F. (2011). Resolving the deposit dilemma: A new DEA Bank Efficiency Model. *Journal of Banking & Finance*, *35*(11), 2801–2810. https://doi.org/10.1016/j.jbankfin.2011.03.007

Isik, I., & Hassan, M. K. (2003). Financial disruption and bank productivity: The 1994 experience of Turkish Banks. *The Quarterly Review of Economics and Finance*, *43*(2), 291–320. https://doi.org/10.1016/s1062-9769(02)00194-1

Lin, T. T., Lee, C.-C., & Chiu, T.-F. (2009). Application of DEA in analyzing a bank’s operating performance. *Expert Systems with Applications*, *36*(5), 8883–8891. https://doi.org/10.1016/j.eswa.2008.11.018

Li, Y. (2020). Analyzing efficiencies of city commercial banks in China: An application of the bootstrapped DEA approach. *Pacific-Basin Finance Journal*, *62*, 101372. https://doi.org/10.1016/j.pacfin.2020.101372

Lee, T.-H., & Chih, S.-H. (2013). Does financial regulation enhance or impede the efficiency of China's listed commercial banks? A Dynamic Perspective. *Emerging Markets Finance and Trade*, *49*(sup4), 132–149. https://doi.org/10.2753/ree1540-496x4905s410

Milenković, N., Radovanov, B., Kalaš, B., & Horvat, A. M. (2022). External two stage DEA analysis of bank efficiency in West Balkan countries. *Sustainability*, *14(2)*, 978. https://doi.org/10.3390/su14020978

Menicucci, E., & Paolucci, G. (2016). The determinants of bank profitability: Empirical evidence from European banking sector. *Journal of Financial Reporting and Accounting*, *14*(1), 86–115. https://doi.org/10.1108/jfra-05-2015-0060

Neves, M. E., Gouveia, M. D., & Proença, C. A. (2020). European Bank's performance and efficiency. *Journal of Risk and Financial Management*, *13*(4), 67. https://doi.org/10.3390/jrfm13040067

Paradi, J. C., Rouatt, S., & Zhu, H. (2011). Two-stage evaluation of Bank branch efficiency using data envelopment analysis. *Omega*, *39*(1), 99–109. https://doi.org/10.1016/j.omega.2010.04.002

Řepková, I. (2014). Efficiency of the Czech banking sector employing the DEA window analysis approach. *Procedia Economics and Finance, 12*, 587–596. https://doi.org/10.1016/s2212-5671(14)00383-9

Řepková, I. (2015). Banking efficiency determinants in the Czech banking sector. *Procedia Economics and Finance*, *23*, 191–196. https://doi.org/10.1016/s2212-5671(15)00367-6

Sealey, C. W., & Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at Depository Financial Institutions. *The Journal of Finance*, *32*(4), 1251–1266. https://doi.org/10.1111/j.1540-6261.1977.tb03324.x

Sufian, F., Kamarudin, F., & Nassir, A. md. (2016). Determinants of efficiency in the Malaysian banking sector: Does bank origins matter? *Intellectual Economics*, *10*(1), 38–54. https://doi.org/10.1016/j.intele.2016.04.002

Schaffnit, C., Rosen, D., & Paradi, J. C. (1997). Best practice analysis of bank branches:stra An application of DEA in a large Canadian Bank. *European Journal of Operational Research*, *98*(2), 269–289. https://doi.org/10.1016/s0377-2217(96)00347-5

Sheytanova, T. (2014) The Accuracy of the Hausman Test in Panel Data: a Monte Carlo Study. (Advanced level thesis). Örebro University, School of Business. Retrieved December 14, 2022, from http://oru.diva-portal.org/smash/get/diva2:805823/FULLTEXT01.pdf

Strange, N., Bogetoft, P., Aalmo, G. O., Talbot, B., Holt, A. H., & Astrup, R. (2021). Applications of DEA and SFA in Benchmarking Studies in Forestry: State-of-the-art and Future Directions. *International Journal of Forest Engineering*, *32*(sup1), 87–96. https://doi.org/10.1080/14942119.2021.1914809

Silva, T. C., Tabak, B. M., Cajueiro, D. O., & Dias, M. V. (2017). A comparison of DEA and SFA using micro- and macro-level perspectives: Efficiency of Chinese local banks. *Physica A: Statistical Mechanics and Its Applications*, *469*, 216–223. https://doi.org/10.1016/j.physa.2016.11.041

Stiroh, K.J. and Rumble, A. (2006), "The dark side of diversification: the case of US financial holding companies", Journal of Banking and Finance, Vol. 30 No. 8, pp. 2131-2161.

Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts, and Urban Transit. *Journal of Productivity Analysis*, *4*(1-2), 183–210. https://doi.org/10.1007/bf01073473

Thoraneenitiyan, N., & Avkiran, N. K. (2009). Measuring the impact of restructuring and country-specific factors on the efficiency of post-crisis East Asian banking systems: Integrating DEA with SFA. *Socio-Economic Planning Sciences*, *43*(4), 240–252. https://doi.org/10.1016/j.seps.2008.12.002

Vodová, P. (2014). Czech commercial banks: Are they liquid enough to finance loan commitments? *Procedia Economics and Finance*, *12*, 752–760. https://doi.org/10.1016/s2212-5671(14)00402-x

Wu, D. D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian Bank. *Expert Systems with Applications*, *31*(1), 108–115. https://doi.org/10.1016/j.eswa.2005.09.034

Wang, K., Huang, W., Wu, J., & Liu, Y.-N. (2014). Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. *Omega*, *44*, 5–20. https://doi.org/10.1016/j.omega.2013.09.005

APPENDICES

Appendix 1: Codes and summary outputs for Efficiency and Tobit estimates

DEA code in RStudio (BCC model)

```
library(Benchmarking)
library(readxl)
File_R <- read_excel("C:/Users/bushra/Downloads/18.11/after/File R.xlsx")
View(File_R)

summary(File_R)
class(File_R)
str(File_R)

x<-with(File_R,cbind(`Fixed Assets`,`Deposits`,`Personnel expenses`,`Other operating
expenses`))
y<-with(File_R,cbind(`Total Loans`,`Other earning assets`,`Off balance sheet items`))


play<-dea(x,y,RTS = 'vrs',ORIENTATION = 'in')
play
(1-eff(play))*x
bcc<-dea(x,y,RTS = 'vrs',ORIENTATION = 'in')
bcc
eff(bcc)

getOption("max.print")
options(max.print = 100000)
data.frame(bcc$eff)
summary(bcc)
```

```
dea.plot(x,y,RTS = 'vrs',ORIENTATION = 'in-out')
dea.plot.frontier(x,y,txt=1:dim(x[1]))

df<-data.frame(bcc$eff)
df
sink('Efficiency data_VRS.xls')
print(df)
sink()
```

Summary output (BCC model)

```
Summary of efficiencies
VRS technology and input orientated efficiency
Number of firms with efficiency ==1 are 102 out of 3952
Mean efficiency: 0.459
---
 Eff range          #        %
 0.1<= E <0.2       1      0.025
 0.2<= E <0.3     443     11.210
 0.3<= E <0.4    1456     36.842
 0.4<= E <0.5     914     23.128
 0.5<= E <0.6     461     11.665
 0.6<= E <0.7     276      6.984
 0.7<= E <0.8     139      3.517
 0.8<= E <0.9      92      2.328
 0.9<= E <1        68      1.721
       E ==1      102      2.581
  Min.   1st Qu.  Median    Mean  3rd Qu.    Max.
0.1879   0.3368   0.4063  0.4592   0.5269  1.0000
```

DEA code in RStudio (CCR model)

```
library(Benchmarking)
```

```
library(readxl)
File_R <- read_excel("C:/Users/bushra/Downloads/18.11/after/File R.xlsx")
View(File_R)

summary(File_R)
class(File_R)
str(File_R)

x<-with(File_R,cbind(`Fixed Assets`,`Deposits`,`Personnel expenses`,`Other operating
expenses`))
y<-with(File_R,cbind(`Total Loans`,`Other earning assets`,`Off balance sheet items`))

play<-dea(x,y,RTS = 'crs',ORIENTATION = 'in')
play
(1-eff(play))*x
ccr<-dea(x,y,RTS = 'crs',ORIENTATION = 'in')
ccr
eff(ccr)

getOption("max.print")
options(max.print = 100000)
data.frame(bcc$eff)
summary(ccr)
dea.plot(x,y,RTS = 'crs',ORIENTATION = 'in-out')
dea.plot.frontier(x,y,txt=1:dim(x[1]))

df<-data.frame(bcc$eff)
df
sink('Efficiency data_CRS.xls')
print(df)
sink()
```

Summary of outputs (CCR model)

Summary of efficiencies

CRS technology and input orientated efficiency

Number of firms with efficiency==1 are 25 out of 3952

Mean efficiency: 0.388

Eff range	#	%
0<= E <0.1	1	0.025
0.1<= E <0.2	9	0.228
0.2<= E <0.3	783	19.813
0.3<= E <0.4	1958	49.545
0.4<= E <0.5	704	17.814
0.5<= E <0.6	208	5.263
0.6<= E <0.7	146	3.694
0.7<= E <0.8	60	1.518
0.8<= E <0.9	43	1.088
0.9<= E <1	15	0.380
E ==1	25	0.633

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.07613	0.30841	0.35483	0.38821	0.41960	1.00000

Tobit regression code

```
library(readxl)

installed.packages('censReg')
library('censReg')

CRS <- read_excel("C:/Users/bushra/OneDrive/Desktop/REGRESSION FILES/CRS.xlsx")
View(CRS)
```

```
VRS <- read_excel("C:/Users/bushra/OneDrive/Desktop/REGRESSION FILES/VRS.xlsx")
View(VRS)

CRS_with_GDP <- read_excel("C:/Users/bushra/OneDrive/Desktop/REGRESSION FILES/CRS
with GDP.xlsx")
View(CRS_with_GDP)

VRS_with_GDP <- read_excel("C:/Users/bushra/OneDrive/Desktop/REGRESSION
FILES/VRS with GDP.xlsx")
View(VRS_with_GDP)

#CRS
R<-censReg(Efficiency ~ SIZE + LOANS + EQUITY + LD,right = 1,data = CRS)
summary(R)

#VRS
S<-censReg(Efficiency ~ SIZE + LOANS + EQUITY + LD,right = 1,data = VRS)
summary(S)

#CRS_with_GDP
K<-censReg(Efficiency ~ SIZE + LOANS + EQUITY + LD + GDP,right = 1,data =
CRS_with_GDP)
summary(K)

#VRS_with_GDP
L<-censReg(Efficiency ~ SIZE + LOANS + EQUITY + LD + GDP,right = 1,data =
VRS_with_GDP)
summary(L)
```

<u>Summary output</u>

1) Dependent variable: TE (CRS)

Call:

censReg(formula = Efficiency ~ SIZE + LOANTA + EQTA + LOANDP,

right = 1, data = CRS)

Observations:

Total	Left-censored	Uncensored	Right-censored
3952	0	3927	25

Coefficients:

	Estimate	Std. error	t value	Pr(> t)
(Intercept)	0.544618	0.015584	34.947	< 2e-16 ***
SIZE	-0.011821	0.001633	-7.240	4.49e-13 ***
LOANTA	-0.707077	0.012538	-56.396	< 2e-16 ***
EQTA	0.058163	0.059759	0.973	0.33
LOANDP	0.426498	0.009422	45.267	< 2e-16 ***
logSigma	-2.394389	0.011308	-211.740	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Newton-Raphson maximisation, 9 iterations

Return code 1: gradient close to zero (gradtol)

Log-likelihood: 3790.271 on 6 Df

2) Dependent variable: TE (VRS)

Call:

censReg(formula = Efficiency ~ SIZE + LOANTA + EQTA + LOANDP,

right = 1, data = VRS)

Observations:

Total	Left-censored	Uncensored	Right-censored
3952	0	3850	102

Coefficients:

	Estimate	Std. error	t value	Pr(> t)
(Intercept)	0.886225	0.023209	38.184	< 2e-16 ***
SIZE	-0.054695	0.002426	-22.541	< 2e-16 ***
LOANTA	-0.774092	0.018858	-41.048	< 2e-16 ***
EQTA	0.442801	0.088795	4.987	6.14e-07 ***
LOANDP	0.473488	0.014443	32.783	< 2e-16 ***
logSigma	-2.001727	0.011491	-174.200	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Newton-Raphson maximisation, 8 iterations

Return code 8: successive function values within relative tolerance limit (reltol)

Log-likelihood: 2091.195 on 6 Df

Appendix 2: Robustness check: Results without the variable SIZE

	(9.1) TE_{CRS}	(9.2) TE_{VRS}
Intercept	0.464***	0.485***
	(0.009)	(0.014)
LOANTA	-0.696***	-0.784***
	(0.012)	(0.019)
EQTA	0.144*	0.810***
	(0.059)	(0.091)
LOANDP	0.405***	0.457***
	(0.009)	(0.013)
Observations	3952	3952
R^2	0.475	0.335

Source: Compiled by author using Microsoft Excel

Note: 1) This table shows the results of OLS regression models for both dependent variables

2) Standard errors are in parenthesis

3) ***p<0.01, **p<0.05, *p<0.10

Appendix 3: Robustness check: Results without the variable LOANTA

	(9.1) TE_{CRS}	(9.2) TE_{VRS}
Intercept	0.476***	0.817***
	(0.021)	(0.027)
SIZE	-0.026***	-0.069***
	(0.002)	(0.003)
EQTA	-0.009	0.350***
	(0.080)	(0.104)
LOANDP	0.114***	0.118***
	(0.009)	(0.012)
Observations	3952	3952
R^2	0.073	0.168

Source: Compiled by author using Microsoft Excel

Note: 1) This table shows the results of OLS regression models for both dependent variables

2) Standard errors are in parenthesis

3) ***p<0.01, **p<0.05, *p<0.10

Appendix 4: Robustness check: Results without the variable EQTA

	(9.1) TE_{CRS}	(9.2) TE_{VRS}
Intercept	0.570***	0.966***
	(0.013)	(0.0190)
SIZE	-0.013***	-0.0577***
	(0.002)	(0.002)
LOANTA	-0.682***	-0.7239***
	(0.012)	(0.017)
LOANDP	0.395***	0.413***
	(0.009)	(0.012)
Observations	3952	3952
R^2	0.483	0.413

Source: Compiled by author using Microsoft Excel

Note: 1) This table shows the results of OLS regression models for both dependent variables

2) Standard errors are in parenthesis

3) ***p<0.01, **p<0.05, *p<0.10

Appendix 5: Link to dataset

The following link contains the raw data collected from BankFocus, the final dataset after data cleaning, as well as two folders, one of which contains the files to run the DEA, and the other one contains files to run the Tobit regression in RStudio. The codes to run these files are provided in Appendix 1.

https://drive.google.com/drive/folders/1PBXgyrr0yFqwPpKA2Vl2Tr9d2EvDZack?usp=share_link

Appendix 6: Non-exclusive licence

A non-exclusive licence for reproduction and publication of a graduation thesis[1]

I, Bushra Faroquei

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis

EFFICIENCY ANALYSIS OF EUROPEAN BANKS: A TWO-STAGE DEA APPROACH

supervised by Simona Ferraro, PhD

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

15.12.2022

[1] *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*