EFFICIENCY IN EUROPEAN BANKING: A RISK PERSPECTIVE

by

Amer Al Bkhetan

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire

April 2020

University of Central Lancashire

STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

Signature of Candidate ______________________________

AMER AL BKHETAN

Type of Award **Doctor of Philosophy**

School **Business School, University of Central Lancashire**

ABSTRACT

This research aims (1) to risk-modify the functional forms applied to estimate technical efficiency in European banking, and (2) to evaluate the sensitivity of estimates to risk and to functional forms' specification.

This research methodology applies the Fourier Flexible functional forms to derive estimates using accounting data from the Bankscope database. For efficiency analysis, profit and cost frontiers are estimated using the Stochastic Frontier Analysis (SFA) in a single-stage estimation approach where the inefficiency term is specified under the time-flexible Battese & Coelli (1995) model.

The past studies tend to overlook a number of significant banking risks, or restrict the estimated functional form, or do both. This research demonstrates that it is paramount to comprehensively accounting for risk in the analysis of efficiency to obtain accurate estimates, and that estimates are very sensitive to the specification of the underlying functional form.

The main limitation to this research related to data availability: data covers 10-year period (2008 – 2018) with 541 only observations. Moreover, risk factors could have been measured using more advanced methodologies (e.g. Merton Approach for measuring the probability of default) which paves the way for further research.

The findings result show that unless risk is adequately accounted for, efficiency estimates tend to be misleading or even erroneous. Findings indicate that European banks can enhance their cost efficiency: (1) by being better capitalized, and (2) by better managing the exposure to trading risk. On the other hand, European banks can enhance their profit efficiency mainly by: (1) maintaining high quality loan portfolios, (2) focusing on interest-based income, (3) relying more on customer deposits as a main source of funding and (4) better managing trading risk exposure to enhance the risk-return payoff. Moreover, results suggest that there is more profit inefficiency than cost inefficiency in European banking overall.

Table of Contents

ABSTRACT *1*

Table of Contents 2

CHAPTER 1: INTRODUCTION *5*

1.2 Background **6**

1.2 Research Aims **7**

1.3 Research Objectives **8**

1.4 Research Contributions: Contribution to Efficiency Literature **9**

CHAPTER 2: LITERATURE REVIEW *13*

2.1 Introduction **14**

2.2 Efficiency Analysis – European Studies **14**

2.3 Efficiency Analysis – International Studies **30**

2.3.1 US Studies 30

2.3.2 Rest of the World Studies 41

2.5 An Extended Summary **50**

2.5.1 European Studies 50

2.5.2 International Studies 53

2.5.2.1 US Studies 53

2.5.2.2 Rest of the World Studies 56

2.6 Brief Summary **57**

CHAPTER 3: METHODOLOGY *60*

3.1 Introduction **61**

3.2 Efficiency Analysis **61**

3.2.1. Production Frontier **61**

3.2.2 Technical and Allocative Efficiencies 65

3.2.3 Bank Production and Cost Functions 68

3.2.4 Model Specification 73

3.2.5 The Cost Function 74

3.2.6 Standard and Alternative Profit Functions 76

3.2.7 Estimating Technical Inefficiency 83

3.2.8 Modelling Inefficiency Effects 86

3.2.9 Technical Inefficiency Test 91

3.2.10 Testing for Stochasticity 91

3.2.11 The Two-Stage and Single-Stage Frontier Estimation 92

3.2.12 Analysing Profit and Cost Efficiencies 93

3.2.13 Bank Inputs and Outputs Specification 94

3.2.13.1 Intermediation and Production approaches 94

3.2.13.2 The importance of accounting for OBS items 98

3.3 Functional Form Specification **99**

3.3.1 The Translog Form 100

3.2.2 The Fourier Flexible Form 101

3.2.3 Regulatory Conditions 104

3.2.4 Structural Test: The Log-likelihood Ratio 106

3.2.5 Frontier Estimation Techniques 108

3.2.6 Non-Parametric Techniques 110

3.2.6.1 Data Envelopment Analysis DEA ... 111
3.2.6.2 Free Disposal Hull FDH ... 115
3.2.7 Parametric Techniques ... 118
3.2.7.1 Distribution Free Approach DFA ... 119
3.2.7.2 Thick Frontier Analysis TFA ... 121
3.2.7.3 Stochastic Frontier Analysis SFA ... 123
3.2.7.4 The Rationale for Choosing SFA ... 128
3.2.8 Maximum-Likelihood Estimation MLE ... 130
3.2.8.1 The Maximum Likelihood Estimation Procedure ... 134

CHAPTER 4: RISK ... 137

4.1 Introduction ... 138

4.2 The Changing Environment of Banking Risks ... 140
4.2.1 Globalization and Financial Crises ... 143
4.2.2 Financial Innovation and Technological Advances ... 144
4.2.3 Securitization ... 148
4.2.4 Interest Rate Volatility ... 148
4.2.5 Financial Regulation: The Evolution of Basel II ... 150

4.3 Major Banking Risks ... 152
4.3.1 Liquidity Risk ... 152
4.3.1.1 Liquidity Risk: Definition ... 152
4.3.1.2 Liquidity Risk: An Industry Perspective ... 153
4.3.2 Credit Risk ... 155
4.3.2.1 Credit Risk: Definition ... 155
4.3.2.2 Credit Risk Appetite CRA ... 156
4.3.3 Credit Risk: A Regulatory Perspective ... 158
4.3.3.1 The Concept of Risk Weighted Assets RWA ... 158
4.3.3.2 The Standardized Approach SA ... 161
4.3.3.3 Credit risk: A Practical Perspective ... 162
4.3.4 Market Risk ... 163
4.2.4.1 Market Risk: Definition ... 163
4.2.4.2 Non-trading Risk ... 164
4.2.4.3 Trading risk ... 165
4.2.4.4 Trading Risk Measurement: A Practical Perspective ... 166
4.2.4.4.1 Stress Testing Model ... 166
4.2.4.4.2 Value at Risk (VaR) ... 167
3.4.5 Insolvency Risk ... 171
3.4.5.1 Insolvency Risk: Definition ... 171
3.4.5.2 Insolvency Risk: Case Studies ... 171
3.4.6 The Integration of Banking Risks in Empirical Analysis ... 173
3.4.6.1 Credit risk: A Research Perspective ... 174
3.4.6.2 Market Trading Risk: A research perspective ... 174
3.4.6.3 Liquidity risk: A Research Perspective ... 175
3.4.6.4 Insolvency risk: A research perspective ... 175
3.4.7 Summary ... 176

CHAPTER 5: EMPIRICAL ANALYSIS ... 179

5.1 Research Sample and Data ... 180
5.1.2 Standardizing data denominations ... 182
5.1.3 Variable Specification 1: Frontier Variables ... 182
5.1.4 Dependent Variables Specification (Profits & Total Costs) ... 185
5.1.5 Inputs and Outputs specification ... 187
5.1.6 Accounting for Equity Capital ... 190
5.1.7 Summary Statistics: A Brief Discussion ... 193
5.1.8 Variable Specification 2: Inefficiency Explanatory Variables ... 198

5.2 Basic Translog Model 'T-Basic' **202**
5.2.1 Testing the Appropriateness of 'T-Basic' 203
5.2.2 Translog Cost function 'T-C' 205
5.2.3 Translog Profit function 'T-P' 207
5.2.4 Fourier Flexible Cost Function 'F-C' 208
5.2.5 Fourier Flexible Profit function 'F-P' 210
5.2.6 Risk-Modified Fourier Flexible Cost Model 'R-F-C' 213
5.2.7 Risk-Modified Fourier Flexible Profit Model 'R-F-P' 214

5.3 Estimating the Preferred Profit and Cost Models **216**

5.4 Empirical Results 1: Cost Function (R-F-C) **220**
5.4.1 Hypothesis Testing 220
5.4.2 Technical Cost Inefficiency Test 221
5.4.3 Stochasticity Test 221
5.4.4 Inefficiency Correlates Test 222
5.4.4.1 Estimating the Cost Function (R-F-C) 223
5.4.5 Determinants of Technical Cost Inefficiency 231

5.5 Empirical Results 2: Profit Function (R-F-P) **237**
5.5.1 Hypothesis Testing 238
5.5.1.1 Technical Profit Inefficiency Test 238
5.5.2 Stochasticity Test 238
5.5.3 Inefficiency Correlates Test 239
5.5.4 Estimating Profit Function (R-F-P) 239
5.5.5 Determinants of Technical Profit Inefficiency 248
5.5.6 Profit Efficiencies by Asset Group 255
5.5.6.1 Correlation Between Efficiencies 257

CHAPTER 7: CONCLUSION* *258

7.1 Summary and Conclusion **259**

7.2 Implications for Policy Makers and Further Research **261**

REFERENCES* *263

CHAPTER 1: INTRODUCTION

1.2 Background

"Banks are of central importance for economic growth, credit allocation, and financial stability" (Molyneux and Wilson, 2007, p 1907). Banks play a key role in allocating financial resources in the economy, and for banks to do this actively they need to be sound and efficient. Operating efficiently is important for banks to achieve growth, maintain competitiveness, increase profitability, and to contribute in to the stability and efficiency of the entire financial system.

An efficient banking system is a socially desirable target to achieve since it reduces the average cost of financial transactions and enhances the society's welfare. With regard to European banking, efficiency has become an important topic given the increased competition the sector has been undergoing even though full integration is yet to be realized. The formulation of the Single European Market, the Second Banking Directive, and the adoption of the Euro for instance have opened new business places for banks to develop, but also intensified competition, squeezing bank margins and prompting banks to improve their cost and profit efficiency (Pastor, 2002).

Banks can operate efficiently by achieving technical efficiency and by take advantage in potential economies. This results in reducing costs and increasing profitability as well. However, risk should not be ignored in this case, and that is why this research attempts to align more closely the analysis of technical efficiency with banking risks. Ignoring risk may yield misleading results and therefore produce unreliable policy implications. For instance, banks could reduce their costs and take excessive risks in the short-run to enhance their profitability. These banks could be erroneously labelled as cost efficient if their risk profiles are not considered in the analysis. In any case, correct and accurate estimates are vital to policy makers.

This observation is confirmed by Girardone et al (2004) who find that "risk and asset quality factors appear to matter in relation to estimation … and these factors should be borne in mind in any future evaluation of the efficiency characteristics of European banking markets" (p 225). Therefore, accounting for risk in analysing and technical efficiency is pivotal. This research confirms this observation and further demonstrates that unless risk is adequately accounted for, efficiency estimates tend to be misleading or even erroneous. This constitutes a key contribution of this research.

The following sections will be discussing the motives behind investigating technical efficiency in European banking.

1.2 Research Aims

This research motives to investigating profit and cost efficiencies in European banking are threefold: (1) incorporating risk into the analysis of efficiency more comprehensively, (2) conducting a European cross-country efficiency analysis, and (3) investigating the correlates of both cost and profit efficiencies in European banking.

Concerning the first motive, risk in this analysis is more comprehensively accounted for by incorporating bank- as well as country-level risks into the specification of cost and profit functional forms. These risks constitute the main banking risks including credit, trading, liquidity, and insolvency risks. Moreover, credit risk is accounted for from two perspectives: an ex ante (using the credit risk appetite measure) and an ex post (using the loan loss provisions ratio). Despite that many authors have stressed the imperativeness of accounting for risk, the vast majority of past efficiency research accounts for risk incomprehensively and at a bank-specific level only in the form of equity capital (to proxy for insolvency risk) and loan loss provisions ratio (to proxy for credit risk or output quality). Mester (1996, p 1026) for instance asserts, “Unless quality and risk are accounted for, one might easily miscalculate a bank’s level of inefficiency”. Drake et al (2006, p 1451) confirms the vitality of incorporating risk factors in into the cost function as “failure to adequately account for risk can have a significant impact on relative efficiency scores”.

As for the second motive, Berger et al (1993b) highlight a significant gap in the literature in that there is a limited cross-country banking efficiency analysis and suggest that “much more research is needed measuring and comparing the efficiency of banks and other financial institutions across international borders” (p 232). Berger et al also argue that the integration of European banking systems would suggest that the most efficient institutions may eventually dominate world markets, thus it is a worthwhile project to conduct European cross-country analysis. This is further echoed by Maudos et al (2002) who find that cross-country efficiency studies are quite limited because “there are few studies that run comparisons at an international level” (p 33). In light of this, it is an interesting project to investigate a European cross-country efficiency research.

With regards to the third motive, investigating efficiency correlates is important because it can provide insights into factors driving banks efficiency so that policy implications can be drawn. Goddard et al (2007, p 1925) stated, “The investigation of factors that are correlated with measured operational efficiency at bank level seems a worthwhile enterprise”. Such an issue is of a particular relevance to European banks given the regulatory transformation it is experiencing and the attempts to achieve further integration in the EU banking system (European Commission, 2007). In addition, Girardone et al (2004) also stress that future research should investigate both cost and profit efficiencies in European banking as banks should not only be evaluated in terms of their ability to use resources effectively to produce a given level products and services (minimize costs), but also their skill at generating maximum revenues from their outputs (maximize profits).

The abovementioned arguments and quotes from past research give a strong rationale to the need to more comprehensively incorporate risk into efficiency analysis, to conduct a European cross-country analysis, and to investigate the correlates of both cost and profit efficiencies in European banking.

1.3 Research Objectives

This research has four main objectives:

1) to risk-modify the profit and cost functional forms applied to estimate technical efficiency in European banking using proxies for: credit, trading, liquidity, and insolvency risks. The central added-value of this research is measuring market and credit risks differently. Market and credit risks will be measured using the VaR model and Merton Approach respectively. These risk variables are comprehensively incorporated in the functional forms at bank- and country-specific levels. To ensure the viability of this objective, tests will be conducted

to investigate whether or not the introduction of these risk variables has a significant impact in terms of improving the functional forms' specification prior to estimating the model.

2) to evaluate the sensitivity of profit and cost efficiency estimates as well as estimates to the introduction of these risk factors, and to the different (flexible Vs. restrictive) Fourier series specifications.

3) to provide European banking cross-country profit and cost efficiency estimates at both bank- and country-specific levels, and also to provide estimates for European banking over 2008 – 2018.

4) to draw implications for policy makers in the European banking industry which will be based on the findings of the two empirical analyses.

1.4 Research Contributions: Contribution to Efficiency Literature

This research contributes to efficiency literature in three aspects. First, the introduction and the specification of Fourier terms is shown to be an important factor for the cost function to satisfy the regulatory conditions. Introducing the Fourier terms is motivated by the observation of Gallant (1982) who finds that Fourier terms significantly improve the model's fit to the scatter of the data (i.e. it being a global approximation). This agrees with Berger and Humphrey (1997, p 9) who observe that the way to improve the performance of parametric approaches "lies in adding more flexibility" to the cost function. This research confirms the observations of Gallant (1982) and Berger and Humphrey (1997) and finds that unless the 1st order Fourier terms are accounted for, the cost model violates the regulatory conditions, particularly in relation to the concavity condition[1]. Furthermore, unlike most efficiency studies, this research follows a careful procedure in specifying the Fourier terms and reveals that the most suitable specification, given the underlying data, is a specification, which comprises of both rescaled input prices and output quantities. In addition, the choice of normalizing input prices was also carefully considered on the basis of satisfying the cost function's

[1] Introducing the second-order Fourier terms seem to overfit the scatter of the underlying data and causes the model to violate the concavity condition as a result.

regulatory conditions. The study finds that the most suitable procedure is to normalize input prices by the price of labour.

Second, this research estimates a common EU-14 profit and cost frontiers and accounts for differences between the countries involved via the incorporation of a range of country-specific macroeconomic variables (including inflation rate, GDP growth, Herfindahl index for loans, and a measure of financial intermediation), dummies to control for EU membership and euro adoption, and risk variables. Past European research is typically restricted to accounting for macroeconomic differences only. Moreover, this contribution is motivated by the observation of the European Central Bank (ECB) in its 2003 banking stability report stating that “aggregate EU data tend to conceal significant differences between countries, as business is still predominantly local” (ECB, 2003, p 8). This in fact highlights the importance of accounting for country-specific characteristics in terms of risk and macroeconomic conditions into the analysis. What is more, it is shown in the literature (as in Berger and Mester, 1997) that accounting for country differences serves the purpose of overcoming the potential shortfall of assuming a single or common frontier for the entire sample. Estimating country-specific or year-specific frontiers on the other hand will create degrees of freedom problems given that the sample used is about 200 observations (McAllister and McManus, 1993), therefore the choice of this research is fitting a single frontier. A number of previous European studies have estimated single frontiers however; they mostly overlooked the issue of accounting for country-specific (i.e. banking systems-specific) differences specifically in terms of their risk profiles.

Third, this study attempts to explain technical efficiency scores by regressing the latter, in a single-stage estimation procedure, against a wide set of explanatory variables reflecting business, funding, and income structures in addition to the set of risk variables. This is motivated by the significant gap in literature observed by Berger and Mester (1997, p 944) where the authors stress the importance for research to further explore the “potential correlates” of profit and cost efficiencies. Risk variables introduced include bank-specific credit risk (that is considered both from an ex ante and ex post perspectives), trading risk, liquidity risk, and insolvency risk. The contribution here lies in the introduction of a wider range of risks into the analysis, most importantly, the credit risk appetite measure CRA (= risk weighted assets RWA / total assets) which was suggested by the Bank of England in its Financial Stability Review of 2003, as it takes an ex ante perspective on credit risk. Credit risk is normally accounted for in the literature form an ex post perspective as represented by the loan quality measure (loan loss provisions / total loans). This research also draws some important policy implications based on the findings of cost and profit efficiency correlates.

Fifth, this research finds that the technical inefficiency term should be specified under a time-flexible specification. This is because estimates show that the time progression as well as the level of technical inefficiency are sensitive to the time-flexibility allowed in the specification of which. The conclusion therefore is not to impose a systematic trend of time progression as in Battese and Coelli (1992), but to allow for greater time flexibility as in Battese and Coelli (1995).

The closest study to this analysis is that by Altunbas et al (2001) where similar methodology to investigate cost efficiency was used (the application of SFA, and the single-stage estimation approach), however the study incorporates Fourier series that are truncated at the 2nd order terms including rescaled values of outputs only, accounts for limited number of risks at a bank level including insolvency risk (equity capital/ total assets), liquidity risk (liquid assets / total assets), an ex post measure of credit risk (non-performing loans/ loans), an ex-ante measure of credit risk (risk weighted assets / total assets), accounts for the funding structure (short-term funding and customer deposits / total funding), accounts for business mix effect (the nominal value of OBS items / total assets) , accounts for a limited number of country-averaged risks including insolvency (equity capital only), liquidity and credit risks, and uses data between 1989 and 1997.

The edge of this research is manifested in using a wider range of bank-specific risk including liquidity, credit (from ex ante and ex post perspectives), insolvency, and trading risks (trading risk is proxied by the ratio of net trading revenue to total revenues). Moreover, this research uses a wider range of country-specific risks including insolvency (equity/ total assets), liquidity, credit, and trading risk. To the best of the researcher's knowledge, there has been no past European research to comprehensively account for banking risks at bank- as well as country- levels, and to carefully consider the specification of the underlying functional form in terms of satisfying the cost function's regulatory conditions. More importantly, there seems to be no previous research that has attempted to test for the impact of CRA, although the concept of risk-weighted assets has existed since Basel I was founded in 1988. What is more, this research incorporates a wider range of macroeconomic variables including EU membership and Euro adoption variables, Herfindahl index for loans, and a proxy for financial intermediation (M2/GDP). Lastly, from a methodological perspective, this research carefully considers the appropriate specification and level of expansion of the Fourier series applied in both empirical chapters using the fulfilment of the concavity condition and the log-likelihood ratio test as the two defining criteria. Such a procedure has never been applied before in the literature.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter provides and extensive review of efficiency literature. The chapter is presented in three main sections: Efficiency Analysis in European, US, and International contexts. Each of these sections is concluded by highlighting the envisaged contribution of this research to the literature respectively.

This review of both cost and profit efficiency literature covers studies that were conducted using European, US, and International sample banks. These studies are displayed and discussed by date beginning with the earliest in each of the subsections. This section of the literature review chapter concludes by extended summaries that sum up the findings of past research accordingly.

2.2 Efficiency Analysis – European Studies

Altunbas et al (2001) examine a sample of European banks in the EU-15 countries over 1989 – 1997 using balance sheet data from Bankscope. The study applies SFA approach to estimate a Fourier Flexible cost function, with the Fourier series including rescaled output quantities only and truncated at the second-order terms.

Altunbas et al (2001) found substantial cost X-inefficiencies ranging between 20 – 25% across European banks. These X-efficiencies were found to considerably vary across European banking markets, bank size, and over time. Technical progress was also found to have a significant impact on total costs, as it seemed to contribute in reducing cost by about 3% over 1989 – 1997. These cost savings were found to increase proportionately with bank size, indicating that Europe's largest banks seemed to benefit most from technical progress.

Cavallo and Rossi (2001) investigate cost X-inefficiencies for a sample of 442 European banks with 2516 observations from France, Germany, Italy, Netherlands, Spain, and the UK over 1992 – 1997. The authors obtain accounting data from Bankscope. The study estimates a Translog cost function and uses Battese and Coelli (1992) inefficiency model in the second-stage estimation. Cavallo and Rossi find significant cost inefficiencies in all of the six countries' banking systems at an average of 15.64%. In terms of size-efficiency relationship, cost efficiency seems to decline as size grows; accordingly, smallest banks were found as most cost efficient as opposed to larger banks. Banking system in Germany was found to be the least cost inefficient, whereas the UK's was found as the most cost inefficient over the study's period.

Vennet (2002) analyse cost and profit efficiencies for financial conglomerates and specialized banks in Europe using SFA approach. Vennet estimates a Translog cost and profit model covering the period 1995 – 1996 and forming a sample with 2375 observations on banks from 17 European countries. The rationale for choosing this time period is that it coincides with the impact of the Second Banking Directive, which has enabled European banks to form financial conglomerates[2] and hold equity stakes in financial and nonfinancial companies, i.e. form universal banks. The study aims at comparing cost and profit efficiencies of European financial conglomerates and specialized banks, and between universal and non-universal banks.

Vennet (2002) considers the level of financial capital as a 'rough' proxy for the risk of default and the risk preference of the bank management. However, the study seems to completely ignore accounting for credit risk as many earlier studies did (in terms of the ratio of problem loans to total loans for instance).

On the other hand, operating costs of financial conglomerates are found to be lower compared to specialized banks due to cost savings realized by exploiting (which were exhausted) and lower funding costs realized by the effects of market power.

Furthermore, universal banks are found to be more profit and cost efficient than their non-universal counterparts. The study attributes this to the better information universal banks (given their equity holdings in the firms they lend to) can access which enables them to mitigate the impact of moral hazard and adverse selection. As results indicate that conglomerates are more revenue and cost efficient than their specialized counterparts.

[2] Financial conglomerates are financial institutions that offer the entire range of financial services. Next to performing the traditional banking operations, they may sell insurance, underwrite securities, and carry out security transactions on behalf of their clients. Universal banks are allowed to hold equity stakes not only in financial but also in nonfinancial companies (Vennet, 2002, p 254).

In terms of efficiency scores, Vennet (2002) finds that an overall cost inefficiency for the entire sample of European banks of about 20%. This is consistent with Berger and Humphrey (1997) who also finds cost inefficiencies to range between 20 – 25% using a large sample drawn from 21 countries. On the other hand, the study also finds average profit inefficiency for European banks at about 30%, suggesting greater inefficiencies on the profit side relative to cost inefficiencies. In terms of the size-efficiency relationship, the study finds little evidence on the correlation between size and efficiency. The only clear evidence found is that smallest universal banks (assets < 500 million ECU) and smallest specialized banks (assets < 100 million ECU) are less profit efficient compared to the corresponding largest banks.

Maudos et al (2002) analyse profit and cost efficiencies using a sample of 3328 observations belonging to 832 banks with minimum asset value of $1bn in ten EU countries over the period 1993 – 1996. The study mainly uses the parametric estimation approach Distribution Free Approach (DFA) suggested by Berger (1993)[3] (which assumes inefficiency as constant or persistent) over time while random errors tend to cancel out over time, therefore it entails no distributional assumptions for the inefficiency term). The study aimed at establishing comparisons between profit and cost efficiencies and also at investigating their potential correlates. Maudos et al use the Translog functional form to describe both cost and alternative profit functions.

In a second stage, the estimated efficiency is regressed against a set of potential correlates. The key finding suggests more bank-specific profit than cost inefficiencies.

As for the profit and cost inefficiency terms' potential correlates, Maudos et al account for several bank-specific characteristics including: bank size, risks, and specialization. Bank-specific risks include: (a) insolvency (or overall) risk proxied by the level of financial capital and the standard deviation of returns on assets ROA (the latter measure was introduced under the assumption that equity capital may not fully capture the amount of risk taken by each bank) and (b) credit risk exposure proxied by the

[3] Maudos et al (2002) also apply the Fixed Effects Model (FEM) and Random Effects Model (REM) to produce efficiency estimates. However, the results of these models are not discussed here because of relative irrelevance to this research.

ratio of loans to total assets. Findings show a non-linear relationship between bank size and profit efficiency, as results show that only medium size banks (up to \$10bn) have a negative and significant correlation with profit efficiency, whereas large banks (\$10bn to \$100bn) and huge banks (over \$100bn) have shown no significant correlation with neither profit nor cost efficiencies.

With regards to bank risks, findings indicate that credit risk (loans/assets) has no significant impact on both technical efficiencies, while the standard deviation of ROA (SD-ROA) was found to be significantly and positively correlated with profit efficiency, and negatively correlated with cost efficiency. This suggests that the riskier the bank is (i.e. the higher the SD-ROA) the more profit efficient but less cost efficient it becomes.

Casu & Girardone (2004) investigate profit and cost efficiencies of large EU banks (assets > € 27 bn) using a sample of around 2363 observations from five EU countries: Germany, France, Italy, Spain and the UK over 1993 – 1997. The parametric estimation approach SFA is applied. The study aimed at examining the potential impact of the introduction of the Single Market Programme SMP in 1992.

Following Berger and Mester (1997), Casu and Girardone apply the alternative profit function, which has the same specification (exogenous variables) as the cost function. The authors use the two-stage estimation approach where the inefficiency term is regressed against a set of correlates in a separate stage to estimating the frontier. Accordingly, the study uses Battese and Collie (1992) specification for the inefficiency model, which assumes that inefficiency change systematically with time.

Concerning cost efficiency results, Casu and Girardone (2004) find average cost efficiency to range between 85.5% to 86.5% (i.e. 14.5% to 13.5% of cost inefficiency) showing around a 1% improvement over time. As for-profit efficiency, average scores were found to be around 87% (or profit inefficiency of 13%) with a significant and dramatic increase of 37% from 53.6% in 1993 up to 90.9% in 1997.

In terms of country-specific efficiency scores (that is, the relative position of each country to the EU-5 frontier), cost efficiency scores show that Italy and the UK were the most and least cost efficient in 1993 (with scores of 89% and 79.9%) and in 1997 (89.7% and 81.1%) respectively. From the profit efficiency perspective, Spain and France where shown to be the most and least profit efficient in 1993 (with scores of 49.9% and 58.2%) and in 1997 (90.6% and 92.6%) respectively.

The authors observe that there is little evidence that cost efficiency levels of the large EU-5 has converged over 1993-97, as country-specific differences seem to play a significant role in the efficiency gap. Such discrepancies in country-specific cost efficiencies can be suggestive of remaining differences in cost bases of the different EU-5 banking systems as a result of considerable differences in taxation, fiscal policies, labour markets flexibility, money and capital markets … etc. On the other hand, profit efficiency estimates show more evidence of narrowing efficiency gap indicating a shift in the focus of EU-5 banks towards becoming more profit-oriented and shareholder value maximizing firms (Casu and Girardone, 2004, p 139). This can support the view that the SMP may have contributed in increasing competition across EU, pressurizing profit margins and therefore reducing the dispersion of profit efficiencies according to the authors.

Girardone et al (2004) estimate cost efficiency in Italian banking between 1993 and 1996. The study mainly aims to investigate the impact of the EU's 1992 Single Market Programme (SMP), explore the potential correlates of efficiency, and to examine the impact of risk of default and asset quality factors (as in Hughes and Mester 1993, McAllister and McManus 1993, Clark 1996, Mester 1996, Berger and Mester 1997, and Altunbas et al, 2000). The introduction of the SMP in 1992 contributed in relaxing the regulatory constraints such as regional branching restrictions, balance sheet portfolio limitations, deposit and loan rate caps, and state ownership. Further, capital markets have become more important in the Italian banking system, as in France, Germany and Spain, which were mainly bank-based financial systems. Against this backdrop of landmark development, the Italian banking sector constituted an interesting area for efficiency research as the authors explain.

Empirically, a stochastic Fourier Flexible cost function is specified. The Fourier terms specified contained rescaled outputs only as input prices were left to be fully described by the Translog terms. The justification for this is that there is little variation in input prices hence the Translog can be more stable (good approximation) near the average of the sample data (Girardone et al, 2004, p 216). The Fourier series, as in Berger and Mester (1997), was truncated at the 3rd order terms. Following the intermediation approach, the authors specify two outputs: loans and other earning assets, while off-balance sheet items were obviously unaccounted for. The study normalizes input prices (labour and financial capital) by the price of real capital (fixed assets). In a second-stage, the estimated inefficiency was regressed (in a logistic regression) against a set of potential correlates.

Findings indicate that overall cost inefficiency levels range between 13 – 15%, indicating that an average Italian bank could have produced the same amount and mix of outputs with 13 – 15% less cost relative to most efficient banks. Results also show that inefficiency seems to decline over time, which can be attributed to the positive impact of deregulation. Also, by plotting X-efficiency scores against size (log of assets), the authors observe a wide dispersion in scores implying that many banks with similar sizes seem to have different cost efficiency levels due to their widely different costs. It follows that X-efficiencies show a declining trend with size, however, no significant correlation between size and cost efficiency was found, despite the overall negative trend found between efficiency and size, implying that larger banks are relatively as cost efficient as their smaller counterparts.

Concerning the correlates of cost inefficiency, results show that cost inefficiency is significantly and positively correlated with the level of none performing loans (NPLs) which indicate that cost efficiency is positively correlated with better credit risk management (credit profile evaluation, loan generation, and loan monitoring). Moreover, cost inefficiency is found to be negatively correlated with capital strength as in Mester (1993 and 1996), that is, banks with better capitalization tend to be more (less) cost efficient (inefficient). Such inverse relationship between capital and cost inefficiency can be explained in two ways. First, banks with low cost inefficiency are more capable of generating profits and therefore retaining more earnings as capital. Second, higher equity capital entails that shareholder have more capital at risk, and therefore have more interest in ensuring that the bank management is operating efficiently, that is to say, cost efficiency seems to be negatively correlated with incentives to moral hazard.

What is more, Girardone et al (2004) further investigate the characteristics of efficient and inefficient banks from different perspectives, including earnings, expenses and balance sheet structure. Findings indicate that efficient banks have higher interest income (in terms of net interest revenue or margin to total assets ratio), better control staff expenses (but no advantage in funding costs), more securities, more equity capital, and better asset quality.

Weill (2004) investigates cost efficiency of European banks over 1992 – 1998 using a sample of 688 banks in five countries: France, Germany, Italy, Spain, and Switzerland. The study aims at crosschecking efficiency results produced by DFA, SFA and DEA using European data as evidence from the US concludes that there is clear inconsistency between efficiency results under parametric and non-parametric techniques (Bauer et al, 1998). Accordingly, Weill (2004) estimates a separate frontier for each country (national frontiers which do not assume that all countries have similar technologies) since the aim is to provide evidence on the consistency of the efficiency results under the different frontier techniques in each of different frameworks (countries). The common frontier methodology, on the other hand, allows for cross-country comparison of efficiency scores but assumes similar production technologies for all countries.

Inputs and outputs are specified according to the intermediation approach, which assumes that the bank transform deposits using labour and capital into loans. Outputs include loans and other earning assets (however off-balance sheet items are ignored), and inputs include labour, physical (real) capital, and borrowed funds (financial capital).

The study finds consistency in cost efficiency results among parametric approaches for all countries (positive and significant correlation), while there seems to be no positive relationship between any parametric approach and DEA. The study also attempts to investigate the relationship between size and cost efficiency. Evidence found is mixer as DEA shows a positive relationship with size, whereas SFA provided no clear correlation between size and efficiency.

To draw further conclusions and cross-check the validity of efficiency estimates under SFA and DEA approaches, Weill (2004) establishes correlations between cost efficiency scores and two profitability indicators: ROA –return on assets– and ROE –return on equity, and a cost effectiveness measure represented by the cost-to-income ratio (all ratios are calculated at the mean for the entire study's period). The following presents the most notable results.

DEA's cost efficiency scores have shown a significant and negative correlation with the two profitability indicators for Switzerland, indicating that banks with higher profitability tend to be less cost efficient. Weill (2004) findings implied by SFA's cost efficiency estimates are in line with those under the standard measure of cost effectiveness, i.e., cost-to-income ratio. Findings show a strong and significantly negative correlation between cost efficiencies and the corresponding results of the cost-to-income ratio for all countries expect for France (where the correlation found was not significant, but negative). In other words, this indicates that greater cost efficiency is associated with lower value of the cost-to-income ratio and vice-a-versa. Such finding indicates to the consistency of cost efficiency estimates under SFA approach, the study concludes.

Casu, Girardone and Molyneux (2004) examine productivity change for large banks in five European countries (France, Germany, Italy, Spain, and the UK) with a panel of 2000 observations drawn from Bankscope over 1994 – 2000 using a Translog cost function. Total factor productivity (TPF) – which measures changes in total output relative to inputs – is decomposed into its core elements: technical efficiency change (TEC) which embodies the effects of pure technical efficiency change TE and scale efficiency change SE, and technological change (TC).

Results indicate that productivity change during 1990s in European banking was mainly driven by technological change (TC). Productivity growth for all counties showed an overall trend of decline especially by the end of sample period, but with significant differences among the 5 banking systems in this respect. More specifically, the authors report a general decline in total costs for the largest Italian and Spanish banks over 1994 – 2000 in contrast to other banking systems. The authors attribute this decline in costs for Spanish and Italian banks to the positive impact of technological change. Furthermore, banks in Spain and Italy experienced a dramatic increase in off-balance sheet activities with the slowest growth in their on-balance sheet assets, suggesting a significant change in the make-up of outputs for Italian and Spanish banks. Results for Germany, France and the UK were inconclusive.

Bos and Kolari (2005) examine profit and cost efficiencies of European and US banks using SFA to estimate cost and profit X-efficiencies. The authors use data on large banks (multibillion dollars in assets) over the period 1995 – 1999 and apply a Translog cost function claiming that Translog and Fourier Flexible functions provide similar results with negligible differences[4]. The intermediation approach is applied in specifying banks' inputs and outputs (which included off-balance sheet items). Risk is accounted for via the ratio of equity to total assets.

In Europe, cost efficiency estimates show that European banks have average score of 0.947 implying that operating costs can further be reduced by 5.3% on average if they were to operate more efficiently. The distribution of efficiency scores was also found to be significantly skewed with the minimum score of 0.064. Further, profit efficiency estimates for European banks were found at a lower average of 0.721 implying that profits can be improved by a further 27.9% on average. The dispersion of profit efficiency results was more substantial as standard deviation of profit efficiencies was found to be as twice that of cost efficiencies which probably indicates to the imperfectly competitive output markets (i.e. the existence of price differentiation). On the other hand, other European studies such as Altunbas et al (2001) found less dispersion in profit efficiencies for large European ranging between 0.20 and 0.25. For US banks, cost efficiency estimates were higher compared to European banks at 0.976 with a lower dispersion (a standard deviation of 0.047).

[4] This can only be credible if the data is concentrated around the mean (i.e. local), whereas most banking data is rather scattered (global). Furthermore, this research demonstrates that the choice between the Translog and Fourier Flexible forms produces very different efficiency estimates.

In general, results show that European banks seem to have lower cost and profit efficiencies than US banks. were US banks are found with smaller dispersion of both profit and cost efficiency sores than European banks. The study also concludes that US and European banks' profit efficiency seems to increase when they expand beyond their home country. Furthermore, evidence from the US and Europe indicates that small banks are less cost and profit efficient than large banks.

Kasman and Yildirim (2006) examine cost and profit efficiencies in commercial banking in eight new member states of the European Union in 2004. The sample used consists of an unbalanced panel data on 190 commercial banks with 977 observations over 1995 – 2002 drawn from Bankscope. The study estimates a single (common) stochastic cost and profit frontiers for the entire sample but accounts for country-specific variables to accommodate for differences in macro-economic and financial sector conditions, which may vary overtime. Such variables included: capital ratio equity/total assets which was used as a proxy for insolvency risk such that lower capital ratio would imply higher insolvency risk resulting in increased cost of borrowed funds), market concentration (Herfindahl) index, inflation, GDP growth, and a measure of banks' financial depth in the economy (proxied by M2/GDP and included to capture differences among the different banking systems in terms of their ability to mediate the lender-borrower relationship). Interestingly, the study uses the single-stage SFA approach to estimate a Fourier Flexible cost and profit functions (that is truncated at the 2nd order terms) where determinants of the inefficiency term are estimated simultaneously with the frontier.

The study found a negative and significant correlation between cost inefficiency and capital ratio implying that well-capitalized banks can provide banking services less costly since they can borrow funds more cheaply. As for the Herfindahl concentration index, the authors explain that if higher concentration is associated with higher cost efficiency (lower costs), it is then attributed to superior management or greater efficiency of the production process. On the other hand, if higher concentration is associated with lower cost efficiency (higher costs), this suggests that higher concentration is the result of market power, which provides less incentive for banks to control costs more carefully.

Furthermore, results showed a positive correlation between market concentration and cost inefficiency, which is consistent with the market power argument. The rate of inflation was also found to cause higher costs given its positive coefficient with cost inefficiency, suggesting that banks incur higher costs under inflationary pressures. On the contrary, GDP growth was found to negatively correlate with cost inefficiency suggesting that with higher growth rate banks tend to bear lower costs. Finally, the measure of overall financial development (M2/GDP) is found to positively contribute to higher cost efficiency suggesting that banks in courtiers with more developed financial intermediation systems tend to operate more cost efficiently.

As for-profit efficiency correlates, findings showed no significant correlation between capital ratio and profit efficiency. The Herfindahl concentration index was found to negatively correlate with inefficiency suggesting that banks are more profit efficient in less concentrated markets. Similarly, inflation was found to negatively impact profit efficiency, implying that inflation contributes to lower profits. On the other hand, the level of financial depth in the economy (M2/GDP) was found to positively impact banks' profit efficiency suggesting that the higher the level of financial intermediation and development the more profit efficient banks are. Unexpectedly, the study found negative correlation between GDP growth and profit efficiency indicating that the higher the GDP growth rate the lower the profits are.

Regarding the size-efficiency relationship (where banks were divided into small banks with less than $1 bn, medium banks with assets greater than $4 bn and less than $5 bn, and large banks with more than $5 bn in total assets), the study found no significant correlation between bank size and profit or cost efficiencies. In terms of efficiency scores, cost inefficiency was found at an average of about 0.20 suggesting that an average bank in new EU member states is incurring 20% more costs than the best-performing or frontier bank. Further, average profit inefficiency was found at around 0.36 suggesting that profits of an average bank are 36% less than those achieved by best-practice banks. This is consistent with earlier evidence from the US in that there is more inefficiency on the profit side than on the cost side.

Altunbas et al (2007) investigate the relationship between capital, risk and efficiency in European banks operating in 15 European countries over 1992 – 2000. The study applies SFA approach to estimate a Translog cost function, which specifies off-balance-sheet items as a third output in addition to loans and securities. Moreover, cost efficiency estimates are regressed against a set of banks- and country-specific variables in a second stage. Bank-specific variables include: (1) credit risk proxied by the ratio of net loans to total assets and the size of loan loss reserves, (2) insolvency risk (proxied by the ratio of equity to total assets), (3) bank size proxied by the log of assets, and (4) liquidity risk proxied by the ratio of liquid assets to customer and short-term deposits.

Country (or banking system) specific variables mainly include: (1) interest rate spreads over 3-year government, and (2) the solvency of the European corporate (non-financial) sector proxied by the ratio of firms' current assets to liabilities. (3) loan loss provisions to total loans, (4) cost to income ratio, (5) return on capital, (6) banking system liquidity. All of the last four ratios were averaged country-wise.

Findings indicate that cost efficiency is negatively and significantly related to the risk and capital ratio. This possibly supports the argument that regulators allow banks that are run in a cost-efficient manner to take more risk (because they can manage it better and accumulate less loan losses accordingly) and therefore hold less capital (become more leveraged), which may reflect greater flexibility towards efficient banks to trade-off capital and risk.

As for country-specific indicators, results show a positive relationship between banking system liquidity and cost inefficiency. However, cost inefficiency was found to negatively correlate with system-specific cost-to-income ratio. The authors attribute such counterintuitive result to the fact that cost efficiency measure was produced from a cost function that links costs to outputs, not income. Therefore, putting this in the context of the positive (negative) correlation found between risk (capital) and cost efficiency, Altunbas et al (2007, p 65) argue that such unexpected result could also be due to the possibility that risk and capital have different impacts on income and outputs (and therefore cost efficiency in this case).

Furthermore, European banks' net lending (net loans to total assets ratio) was found to positively correlate with cost efficiency, indicating that efficient banks are more successful in managing and expanding their loan business. Cost inefficiency was also found to positively correlate with the level of bank capital (represented by the ratio of equity to total assets), suggesting that inefficient banks tend to hold more capital. In terms of size-efficiency relationship, the authors find evidences of a positive correlation between asset size and cost inefficiency, suggesting that larger banks tend to be less efficient than their smaller counterparts. Finally, the study finds mixed evidence on the relationship between liquidity and cost efficiency.

Barros et al (2007) analyse profit and cost efficiencies of European commercial banks using a sample of 1384 observations over the period between 1993 and 2001. The aim is to investigate the determinants of best and worst European domestic and foreign banks' performance. Best (worst) performing banks are defined as those with profit and cost efficiencies in the upper (lower) quartiles of the distribution of estimated efficiency (Barros et al, 2007, p 2198). To this end, the study estimates a Fourier Flexible cost and profit functions to produce efficiency estimates. A logit regression is then applied to examine efficiency correlates in order to investigate factors explaining the probability of banks becoming best or worst performers, however the authors seem to ignore risk completely. Efficiency correlates investigated in this respect include bank location, balance sheet structure and bank ownership.

Results indicate that the probability of being a worst performer is in Denmark, Germany, Portugal or Sweden. However, this probability increases if the bank operates in Finland, Luxembourg or the Netherlands. Competition levels can offer some explanation to these results according to the authors. For Germany and Denmark, the study observes an increased competition in their banking markets, which may have prompted banks to operate more efficiently to secure their market shares. Similar logic can be drawn on Finland's situation where banking markets are more concentrated.

However, this argument does not precisely extend to the case of the Netherlands and Luxembourg which have very competitive markets. The authors acknowledge that the reason why banks in these markets are more likely to be worst performers is related to their profit and cost efficiencies, as they seem to have achieved the lowest efficiency scores.

With regard to the impact of ownership structure, domestic banks were found to be more cost and profit efficient, hence less likely to be labelled as worst performers as opposed to foreign banks. In terms of factors related to banks size and business structure, findings indicate that smaller and loan-intensive banks are unlikely to be among the worst performers compared to bigger and more diversified banks. Nevertheless, it can be argued that these results should probably be treated with caution, as risk –at least in the form of loan losses– was not accounted for in the analysis. Therefore, findings might be biased in favour of riskier banks as earlier research have found that the diversification of risk helps in reducing average costs as it allows for greater exploitation of technical efficiencies (Drake and Hall, 2003).

All in all, the study of Barros et al (2007) provides some interesting results in that there is strong evidence that competitive markets seem to induce banks to become more profit and cost efficient. This has implications for the EU policymakers in that competition should be advocated as this can promote greater efficiency in the banking market EU-wide. This may encourage the provision of more diverse, better-quality, and less-costly products and services by avoiding waste in resources and by fully exploiting under-utilized capabilities.

Iannotta et al (2007) explore the impact of ownership, risk and performance factors on the profitability (operating profits to total earning assets) of 181 largest European banks (forming 1674 observations) from 15 European counties between 1999 and 2004. The study applies OLS regression (as opposed to the frontier estimation) methodology and regresses profitability against a set of banks- and country-specific controlled variables. The study considers several proxies for risk, amongst which are: (1) the ratio of liquid assets to total assets to control for liquidity risk, (2) the ratio of loan loss provisions to total loans to control for credit risk and output quality, (3) the volatility (standard deviation) of asset returns (SD-ROA) to proxy for the bank's insolvency (overall) risk.

Confirming McAllister and McManus (1993), Iannotta et al (2007) find that larger banks can better exploit the advantages of diversification to reduce liquidity risk hence enjoying lower funding costs (and hence higher profitability) compared to smaller banks. The authors observe that this size-risk relationship is partly driven by the too-big-to-fail argument which implies an implicit bail out guarantee that eventually translates into lower costs, which on the other hand encourages larger banks to invest in riskier assets. The study points out that privately-owned bank outperforms mutual and public banks in terms of profitability despite that they have cost advantages. Moreover, public banks were also found to have the lowest loan quality and highest insolvency risk.

On the contrary, mutual banks were found to exhibit the better loan quality and the lower asset risk compared to private and public banks. Ownership concentration was found to be strongly correlated with higher loan quality, lower asset and insolvency risks such that banks with dispersed ownership were found to run higher operational costs per euro of earning assets – a finding that is probably consistent with the implications of agency theory. Given the revealed poor performance of public banks (as they run higher insolvency risk and have lower loan quality), Iannotta et al (2007) recommend that European banking regulators need to abolish any implicit or explicit guarantees to protect government-owned banks and to inform capital markets that the bailout policy of too-big-to-fail is redundant. This way, the authors argue, public banks will be motivated to operate more profitably.

Yildirim et al (2007) investigate cost and profit efficiencies of banks in 12 transition economies in Europe between 1993 and 2000. The aim is to investigate the impact of ownership structure on bank efficiency. The study employs a two-stage estimation approach and uses SFA approach to estimate a Translog cost and alternative profit functions. In the second stage, inefficiencies are regressed against a set of potential determinants. A sample of 325 banks with 2042 observations (drawn from Bankscope) forming an unbalanced panel dataset is used.

Consistent with earlier international and European studies (as in Berger and Humphrey, 1997 and Vennet, 2002), the study finds greater inefficiencies on the profit side than on the cost side. Specifically, the study reports average cost efficiency for the entire sample banks of 77%, whereas (alternative) profit efficiency was found at about 68%.

The second-stage regression analysis showed that higher efficiency levels are associated with well-capitalized banks. This is confirmed by the findings of this research as will be seen in the first empirical chapter. On the other hand, the degree of competition was found to negatively impact profit efficiency whilst positively impact cost efficiency.

Pasiouras (2008) investigates the technical efficiency of the Greek commercial banking industry over 2008 - 2018 following DEA methodology and then conducts a Tobit regression of the inefficiency term against a set of determinants. The study finds that accounting for loan loss provisions (a proxy for credit risk considered) increases efficiency scores. This probably is a result of well-managed loan portfolio in the sense that efficient banks are more capable of managing credit risk more effectively, i.e. maintaining low levels of loan loss provisions. This can further be explained by the positive relationship the author finds between the level of loan activities and efficiency, suggesting that efficient banks are relatively more capable of lending less costly and therefore offer more competitive loan rates, and ultimately expand their loan books and gain greater market share.

On the other hand, the inclusion of off-balance sheet items (OBS) into the analysis as an additional output seems to have no significant impact on efficiency scores. In contrary to the latter finding, this research finds that OBS has a very significant impact on technical efficiency as will be shown later in the first empirical chapter. Interestingly, the study finds that efficiency is positively correlated with the bank's capitalization which is confirmed by the findings of this research as well despite using SFA instead. The argument here is that equity capital represents the capital at risk for shareholders, thus with higher level of equity capitalization, shareholders have more incentive to monitor the bank management and ensure that it is operating efficiently (Pasiouras, 2008, p 313).

Fang et al (2011), examine the cost and profit efficiencies of banking sectors in six transition countries in South-Eastern Europe over the period 1998 – 2008. Fang et al (2008) used SFA (Stochastic Frontier Analysis) approach. Results show that average cost efficiency of bank in these countries is around 69% while average profit efficiency is around 54%. The study regresses determinants of bank efficiency and finds that foreign banks enjoy greater profit efficiency yet lower cost efficiency. Moreover, government-owned banks are associated with lower profit efficiency compared to domestic private banks. The study also finds a strong relationship between market power a bank can have and profit and cost efficiencies. Methodologically, the study uses the Translog form and estimate both cost and profit efficiencies using the Stochastic Frontier Analysis approach SFA which has the advantage of disentangling the error term into stochastic noise and inefficiency. Moreover, the specification of the functional of Fang et al (2011) uses proxies for time, region, GDP growth, and inflation whereas the study completely ignores banking risks. This gap in the literature was clearly highlighted by Fang et al (2011, 517) "Although this article contributes to a better understanding of the determinants of bank efficiency, it is important to highlight that advances in financial sector development will also depend on other *critical factors* that need future research attention". This statement indeed confirms the importance of introducing risk into the analysis of efficiency in banking, a key prospective contribution to this research.

Now, having reviewed EU-related literature, the next section will review the part of efficiency literature that is US-related. This will be followed by international studies and a summary that gives the gist of all of these sections.

2.3 Efficiency Analysis – International Studies

2.3.1 US Studies

Berger et al (1993a) investigate X-efficiency of US commercial banks over the period of 1984 – 1989 using the DFA estimation approach. The study aims primarily at analysing profit efficiency correlates. The authors explain that estimating the profit function has a key advantage in that it allows for detecting inefficiencies not only on the output side (revenue effects) but on the input side (cost effects) of the production process as well. Revenue effects are attributed to producing the incorrect levels or combinations of outputs, and cost effects are attributed to deploying the incorrect levels or mixes of inputs. The study specifies the profit function according to the Fuss function which is an improved functional form of the Translog function. It has an advantage over the Translog in allowing for negative profits and zero values for fixed netputs with the restriction of linear price homogeneity maintained.

Findings indicate that technical inefficiencies dominate allocative efficiencies, suggesting that US banks seem to be less efficient in producing given mixes and levels of outputs using given levels of inputs and input prices than in allocating inputs and choosing output mixes to produce. The other key finding is that there are more output inefficiencies stemming from deficient output revenues than input inefficiencies caused by excessive input costs.

The study also involved regressing profit inefficiency results (expressed as the ratio of profit inefficiency/total assets) against a set of factors including bank risk and size. Risk is accounted for in terms of the standard deviation of the return on assets (SD-ROA). Findings indicate that there is a significant correlation between profit inefficiency and bank risk, suggesting that efficient banks may take fewer risks or be more effective in managing their risks (through diversifying risks) so as to protect their profitable business lines. Finally, in terms of the impact of size (represented by the log of assets), Berger et al find a strong and significant negative correlation between size and profit inefficiency. The authors find the correlation coefficient of this size-efficiency relationship so large suggesting that this is more likely to be a real event as the direction and the significance of this correlation persists even when the data is segmented into different asset groups. This led to the conclusion that larger banks are more profit efficient than smaller banks.

Most importantly, Berger et al (1993a), as later confirmed by Mester (1996) and Mitchell and Onvural (1996), provide evidence that there are more costs to be saved by being technically efficient than by fully exploiting. This suggests that greater cost savings can be made through X-efficiencies in terms of better resource and revenue management, the application of new technologies etc, compared to the potential cost savings resulting from expanding output scale to reduce average cost. Particularly, Berger et al (1993a) found that X-inefficiencies account for 20% or more of banking costs (with technical dominating allocative inefficiencies), whereas make up 5% or less of total costs.

The Federal Reserve Bank of Kansas published a banking efficiency study conducted by Spong et al (1995) in which a risk-modified Translog cost model was applied to produce cost efficiency estimates. The study incorporates the ratio of total loans/ total assets (loan-to-asset ratio) and equity capital / total assets (capital buffer ratio) ratios in the model as proxies for credit and insolvency risk respectively. The study also accounts for output quality by including the ratio of net loan losses/ total loans. Results indicate that banks tend to be more cost efficient if they take more credit risk, i.e. higher loan-to-assets ratio, as opposed to their least efficient counterparts. Moreover, cost efficient banks seem to hold higher levels of capital compared to the less efficient banks. Consequently, there is little support for the argument that efficient banks may have achieved their enhanced efficiency by allocating fewer resources to manage and control risk which probably is suggestive of a positive relationship between risk management and cost efficiency.

Clark (1996) examines X-inefficiencies in US banks using a panel data on 110 publicly-traded commercial banks between 1988 and 1991. The study applies a Translog cost function to produce efficiency estimates. Using the standard deviation of ROA 'returns on assets' as a proxy for risk, Clark finds that X-inefficiency tend to be overestimated if risk is not accounted for. Specifically, X-inefficiencies were found to dramatically drop from 9% to 3% as a result.

Berger and Mester (1997) investigate cost and 'alternative' profit efficiencies in US banking using a sample of about 6000 US commercial banks over 6 years period between 1990 and 1995. The two-stage estimation approach is applied to Fourier Flexible functional forms which were estimated using SFA methodology and followed by a second-stage regression to investigate potential correlates of the estimated inefficiencies.

The study asserts that it is vital to account for output quality when estimating technical efficiencies for the fact that banking data does not fully capture the heterogeneity in bank output[5] which will ultimately affect the credibility of the measured inefficiencies (Berger and Mester, 1997, p 908). To this end, the study uses the proportion of nonperforming loans to total loans (NPL) as a proxy for output quality. However, Berger and Mester draw attention to the issue of how exogenous NPL is to the cost and profit models[6] so as to be correctly included as an independent variable. Therefore, they distinguish between two factors driving the NPL ratio: '*bad luck*' as an *exogenous* factor that is caused by negative economic conditions such as recessions, and '*bad management'* as an *endogenous* factor caused by deficiencies in managing the loan portfolio such as allocating insufficient resources –or '*skimping*' – to assess and originate loans to cut down short-term expenses.

The study solves the exogeneity problem in the frontier models by using the ratio of nonperforming loans to total loans as a state average or state-specific which allows for negative economic shocks to affect banks while being exogenous to individual banks (Berger and Mester, 1997, p 909). On the other hand, bank-specific NPL ratios (attributed to bad management) were included in the second-stage efficiency regression as explanatory variables.

Furthermore, the authors emphasize the significance of accounting for financial capital in the cost and profit functions. The argument here is that these functions assume risk neutrality in banks unless equity capital is introduced such that, risky (risk-averse) banks would be expected to hold relatively lower (higher) level of capital. Berger and Mester (1997, p 909) stress that "if financial capital is ignored, the efficiency of these banks would be mismeasured, even though they behave optimally given their risk preferences". Accordingly, financial capital is accounted for such that outputs and the dependent variable (profits or costs) are normalized by which to accommodate for banks risk preferences and to reduce the effect of size-bias.

Empirically, the study reports higher levels of profit than cost inefficiencies. Mean cost

[5] Commercial loans for instance may differ in size, collateral quality, repayment schedules, risk ...etc. these differences will ultimately impact bank cost and profits, hence ignoring output quality would produce incorrect measures of cost and profit efficiencies.
[6] I.e. how exogenous NPL is to the other independent variables defining the frontiers, as the endogeneity of NPL (due to managerial inefficiency or bad management) would distort technical inefficiency estimates.

efficiency was found at 0.86, suggesting that about 14% of costs are wasted by an average US bank compared to the best-performing banks, while average profit inefficiency was found at around 50% suggesting that about half of an average US bank's profits were lost due to operational inefficiency. Cost efficiency estimates were found to be more tightly distributed (with standard deviation of 6.2 percentage points) while profit efficiencies were found to be quite dispersed (with standard deviation of 20 percentage points). This implies that the earnings of many US banks were considerably dispersed around the average figure compared to the dispersion of their costs.

In terms of efficiency correlates, Berger and Mester examine 25 explanatory variables that mainly include bank size, funding structure (the ratio of purchased funds to total assets), market power (approximated by the HERF index), and risk (approximated by the standard deviation of the bank's annual return on assets SD-ROA and loan-to-asset ratio). As for the size-effect, the study found that larger banks are more cost efficient but less profit efficient compared to their smaller counterparts, that is, as banks grow larger, they become more effective in controlling cost but find it harder to create revenues efficiently. Furthermore, the study finds no significantly positive correlation between profit and cost efficiencies.

In relation to the funding structure, the study reports significant and negative correlation between the ratio of purchased funds to total assets and profit efficiency, suggesting that profit efficient banks would tend to relay more on core deposits and use less purchased funds. As for market concentration, the study finds that market power is negatively related to cost efficiency yet positively correlated with profit efficiency.

As for risk, the study found that riskier banks (i.e. those with higher SD-ROA or more volatile returns) tend to be less profit and cost efficient. The authors attribute this negative relationship between credit risk and technical efficiency to endogenous reasons or bad management. This implies that bad managers are poor at-risk management in terms of taking on poor risk-return trade-offs which involve higher risk levels yet are associated with considerable costs (bad loans) and lower expected returns. This is in contrast to the notion that riskier banks can be expected to be more profit efficient as they trade risk with higher returns.

All in all, the 25 potential correlates were able to explain about 7% of cost efficiency scores' variance and about 35% of profit efficiency scores only. The authors partly attribute this relatively low degree of the model's explanatory power to potential efficiency correlates that are unaccounted for, including risk other factors.

Berger & De Young (1997) specify a Fourier Flexible cost model (that is truncated at the 3rd order terms) and estimate it with SFA approach using US commercial banking data over the period between 1989 and 1994. The frontier function was estimated separately for each year to produce year-wise frontiers under the assumption that an efficient bank in one year may not continue to be so across time.

The study mainly aims at (1) measuring cost efficiency and (2) investigating sources of NPL (non-performing loans) by regressing the latter against a set of variables including the estimated cost efficiency to test the effects of several hypothesis including: bad luck (as proxied by GDP growth), bad management in which loan quality is driven by internal factors (this is proxied by the level of estimated cost efficiency), risk management skimping in which higher loan quality is traded for short-run expenditure savings (this is proxied by the level of financial capital allocated to absorb loan losses), and moral hazard (that is reflected by the nature of the relationship between capital ratio and nonperforming loans).

First, the study finds average cost efficiency over the entire sample period at 0.92, indicating that the average bank incurs about 8% more costs than a best-practice bank. Further, Berger and De Young (1997) test whether specifying a Translog cost function would alter the results. They found cost inefficiency estimates to significantly increase as much as twice the estimate under the Fourier Flexible. The authors attribute this to the global flexibility of the Fourier Flexible specification which enables the frontier to better fit the data compared to the more restricted Translog specification.

Second, the study finds support for both bad management and skimping hypotheses. At the industry level, evidence suggests that the bad management hypothesis seems to dominate the skimping hypothesis (as implied by the corresponding variable's coefficients). This implies that increases in cost *inefficiency* are generally the cause of increases in nonperforming loan levels. Put differently, the deterioration in the quality of loan portfolios, industry-wise, is a result of bad management practices in terms of incurring excess expenditures, poor loan underwriting, and monitoring practices that ultimately lead to increased nonperforming loans.

At a bank level, findings seem to favour the skimping hypothesis, suggesting that increases in cost *inefficiency* are associated with decreases in loan quality such that banks would purposefully trade short-term cost cuts for lower loan quality in the long-term. Furthermore, the study also finds support for the moral hazard hypothesis as suggested by the negative relationship between capital strength and the level of nonperforming loans. In other words, thinly capitalized banks (i.e. those with low capital ratios) would tend to hold loan portfolios with lower qualities (i.e. riskier loans) due to moral hazard. In line with this, Pastor and Serrano (2005) explain that over the short term, the relationship between problem loans and cost efficiency can become misleadingly positive when banks cut expenses on credit assessment and control. The current US mortgage crisis and the associated global credit crunch is a prime example of such case. This reasserts the need for incorporating risk more comprehensively in analyzing bank performance as this research attempts to achieve.

Berger and De Young (1997) conclude that each of these effects (i.e. bad luck, bad management, skimping, and moral hazard) has relatively small impact on banks at an industry level, however their effect is more substantial at a bank-specific level. Therefore, Berger and De Young (1997) assert that the effects of these factors have to correctly be accounted for in profit and cost functions.

For instance, controlling for 'bad luck' – which increases the level of nonperforming loans (hence increase costs or reduce profits) for external, not managerial inefficiency, causes – in the frontier function as a control variable is vital as this would statistically avoid counting these extra costs as inefficiency. This is because the function's residual, which technical inefficiency is part of, will not incorporate the effects of bad luck once it is accounted for in the profit and cost functions as a control variable.

On the other hand, given that the effects of bad management and skimping are the result of management inefficiencies (internal causes), accounting for these factors in the cost or profit function as control variables will artificially increase measured efficiency. This is due to erroneously removing the part of cost inefficiencies (or revenue deficiencies). Consequently, it is important to realize the exogenous and endogenous nature of these different factors and control for exogenous events (bad luck) in the frontier function while controlling for endogenous factors (bad management, skimping, and moral hazard) as determinants of cost or profit inefficiency.

Clark and Siems (2002) examine the impact of accounting for off-balance sheet activities (OBS) on profit and cost efficiency estimates using a sample of US banks over 6 years period (1992 – 1997). They utilize SFA methodology to estimate Translog cost and profit functions. The estimation was conducted with OBS activities specified as an output forming the 'full model' and without OBS activities forming the 'restricted model'. The study uses the volume of net non-interest income (NII) as a proxy for the impact of OBS activities. Under the full model, the study finds mean cost efficiency at 0.8716 suggesting that an average bank incurs about 14% more costs than those of the best-practice banks. On the other hand, mean profit efficiency was found at 0.6462 suggesting that profits of an average bank are approximately 35% below the profits achieved by best-practice banks.

Furthermore, when comparing cost efficiency estimates under the two models, the study finds that cost efficiency becomes higher once OBS activities are accounted for, suggesting that cost efficiency estimates could be biased downwards if OBS activities are not specified. On the other hand, profit efficiency estimates experienced no significant change when OBS items are specified in the profit function. However, the study contrasts these results with earlier efficiency studies, which use 1980s data and ignore OBS activities, and finds that profit efficiency estimates were around 20% to 30% higher once OBS items are included.

In a second stage, the study regresses efficiency estimates against bank size, the ratio (or mix) of off- to on- balance-sheet items, and four OBS variables: derivatives, and loan commitments, lines of credit, and credit commitment activities. Results suggest that neither the relative size of off- to on-balance-sheet activities nor the bank size has any statistically significant correlation with either cost or profit efficiencies. Nonetheless, findings indicate that loan guarantees, credit commitments, and lines of credit have a positive relationship with both cost and profit efficiencies. Derivatives activities, however, show a negative relationship with both profit and cost efficiencies. The study further notes that the correlation of these OBS activities with cost efficiency seem to be much stronger compared with profit efficiency.

To sum up, Clark and Siems (2002) conclude that OBS activities should be accounted for in measuring cost and profit efficiencies of banking organizations as estimates can found to become biased otherwise. Specifically, there is strong evidence showing that cost efficiency tends to be underestimated if OBS activities are not specified in the cost function.

Akhigbe and McNulty (2003) investigate profit efficiency of US banks by mainly focusing on small banks with total assets under $500 mil over 1990 – 1996. To establish comparisons between large and small banks, the study assumes a single frontier (i.e. both types of banks are assumed to use the same production technology) that is estimated under SFA approach (were deviations from the frontier are attributed to inefficiencies and random effects). The study uses the concept of the alternative profit function and applies a Fourier Flexible functional form, where Fourier terms include rescaled output quantities only and truncated at the 3rd level. Estimated profit inefficiency is regressed against a set of potential correlates in a second stage. The study defines banks' outputs according to the production approach where deposits are considered as a third output besides loans and fee-based financial services.

As in Berger and Mester (1997), the authors find that small banks are more profit efficient than large banks. Specifically, profit efficiency was found to improve over the study period which was mainly attributed to improving loan quality. Also, average profit efficiency scores for small and large banks were around 0.75 and 0.72 respectively. The study refers this result to several reasons, including: funding cost, loan quality, market concentration, and bank size.

Akhigbe and McNulty (2003) observe that small banks seem to have better access to retail deposits which gives them a funding cost advantage over large banks since large deposits are more expensive than retail deposits (Akhigbe and McNulty, 2003, p 322). This is besides the fact that demand deposits are a source of fee-based income. Furthermore, given the study's time frame, small banks were found to enjoy higher quality in their loan portfolios compared to larger banks (that is lower non-performing loans ratio to total loans). Also, profit efficiency was found to positively correlate with market concentration, and smaller banks are usually found to operate in rather concentrated markets (as suggested by the Herfindahl index), according to the study. Finally, the study finds that, for banks under \$500 mil, profit efficiency improves as size increases.

In a later study by Akhigbe and McNulty (2005), the authors investigate profit efficiency for small (under \$100 million in assets), medium (above \$100 million and less than \$1 billion), and large US commercial banks (with assets above \$1 billion) over the period of 1995 – 2001. The main aim is to explore potential profit efficiency correlates for each of these asset group banks.

To this end, the study estimates three different frontiers for each of the asset group banks. The reason for this procedure, according to the authors, is that banks are of considerably different sizes hence they are likely to use different business models (production technologies). Accordingly, the resulting profit efficiency estimates are believed to have maximum flexibility (by avoiding biased estimates which could be implied by the common frontier in this case). The authors apply the alternative profit function specified under the Fourier Flexible form that is truncated at the 3rd order terms as in Berger and Mester (1997). In terms of risk, the study accounts for the bank's financial risk by including the level of equity capital as a control variable in the profit functional form. This is to control for the potential impact on cost of funds, and hence on profits, due to different levels of equity capital held.

The study recognizes that equity capital is a rough way to control for risk, but takes no further steps in this regard. Further, the study applies the two-stage estimation approach whereby the efficient frontier is firstly estimated and profit efficiency estimates are regressed against a set of determinants in a second stage (using Tobit regression in this case). Some of these determinants included were: capital ratio (equity/assets), nonperforming loan ratio (included on a country- and bank-specific levels), deposit market concentration index (Herfindahl index), and an income structure proxy represented by the size of fee revenues.

As for-profit efficiency scores, results reveal a positive size-efficiency relationship as large banks are, overall, found to be the most profit efficient compared to small and medium size banks. Specifically, average profit efficiency scores for small, medium, and large banks were found at 0.752, 0.823, and 0.856 respectively.

As for-profit efficiency correlates, results indicate that profit efficiency is significantly and negatively correlated with capital ratio for medium and large size banks. This implies that the more profit efficient banks within these size groups use less equity (more leverage) than other banks in the same size group. A negative correlation is also found between bank-specific nonperforming loans level and profit efficiency. This relationship was found to most significantly impact small bank's efficiency.

Differences in fee revenues are found with significant and positive impact on profit efficiency for small and medium size banks, but not for large banks, suggesting that it is a significant source of profitability for banks in these two size groups. Because differences in fee revenues showed no significant impact on large bank's profitability, this does not mean that fee revenue is an unimportant source of profitability for large banks; rather, it suggests that all large banks in the sample seem to considerably depend on fee revenue to the extent that differences in fee revenues showed no significant impact on profit efficiency as the authors explain. Deposit market concentration index showed, as in Berger and Mester (1997) and many other studies, a positive and significant correlation with profit efficiency of US banks indicating that, banks can be more profit efficient (i.e. claim more profitable) in concentrated deposit markets where they can exert some market power.

Lastly, the study found that small US banks can significantly enhance their profit efficiency by maintaining high asset quality (low default rates), depend more on fee income, and operate in a concentrated deposit market. On the other hand, large banks can significantly enhance their profit efficiency by primarily using more leverage (depend less on equity).

2.3.2 Rest of the World Studies

Allen and Rai (1996) use a sample of 194 international banks operating in 15 different countries (including the US) to investigate cost efficiency using SFA approach to estimate a Translog cost function. Banks were divided into two groups: universal banks (those offering a variety of financial services such as loans, deposits, insurance, securities investment, real states…etc) and non-universal banks (those with functionally-separated traditional and non-traditional banking activities). The study found that universal banks are more efficient than non-universal banks, which offer selected financial services. This could be attributed to the diversification effect, which reduces the overall risk of the bank's balance sheet and translates into less equity capital to be held. This in turn reduces costs and enhances cost efficiency accordingly. Large banks were found with more substantial cost inefficiency of 27.5% compared to small banks, which their input-mix inefficiency averaged around 15% only. At a country level, the authors found that France, Italy, the UK, and the US seemed to host the least cost-efficient banks, whereas banks in Japan, Australia, Austria, Germany, Denmark, Sweden, and Canada are, on average, the most efficient in the world.

On the other hand, the work of Berger and Humphrey (1997) takes an international perspective and surveys 130 efficiency studies applying different parametric and non-parametric estimation approaches that were conducted in 21 countries. Berger and Humphrey mainly found that both types of approaches yield almost similar estimates but produce different rankings for individual banks, and that non-parametric methods generally yield slightly lower mean efficiency estimates but with larger dispersion relative to parametric estimates. More specifically, the study mainly concluded that the overall average level of efficiency (inefficiency) in banking is around 77% (23%) of total bank costs.

Altunbas et al (2000) examine X-inefficiencies for 139 Japanese banks (with 533 observations) over 1993 – 1996. A Fourier-flexible cost function is specified and estimated using SFA approach. Three outputs are considered: total loans, total securities, and total off-balance sheet items (in nominal values). The authors account for the impact of output quality and risk factors on costs. Output quality (credit risk) is accounted for by the ratio of non-performing loans to total loans. Concerning risk factors, the ratio of liquid assets to total assets (to proxy for liquidity risk) and the level of financial capital (to proxy for insolvency risk) were incorporated as in Hughes and Mester (1993) and Mester (1996). These risk factors were considered as endogenous or bank-specific variables. The authors' endogeneity argument stems from that of Berger and De Young's (1997) in that the ratio of non-performing loans is a function of the bank's efficiency and effectiveness in underwriting and monitoring loans. Likewise, liquidity risk is also considered as an endogenous variable since efficient managers would tend to hold low but sufficient levels of liquid assets, whereas inefficient managers would tend to hold excessive levels of liquid assets.

Findings indicate that mean level of X-inefficiency is between 5% and 7%. This is distinguishable from US evidence in Berger and Humphrey (1997) where X-inefficiencies were found about 23% (within the range of 20 – 25%) of total costs. Altunbas et al also observe that X-inefficiencies are quite sensitive to risk and quality factors as estimates were quite different when these factors were discounted. The authors also find evidence of declining costs due to technical change over 1993 – 1996.

Sathye (2001) investigates X-efficiencies (or overall efficiency) in Australian banking system using a cross-sectional data for 29 domestic and foreign banks in 1996 using DEA methodology. The study finds the overall efficiency[7] of Australian banks to be considerably low in comparison with European and US banks. Specifically, technical inefficiency – that is waste in inputs incurred in producing outputs– was found as the main component driving the overall efficiency results relative to allocative inefficiencies

[7] Overall technical efficiency = Pure technical efficiency * Scale efficiency. Drake and Hall (2003, p. 899) explains that scale efficiency arises "because the firm is at an input-output combination that differs from the equivalent constant returns to scale situation", whereas "pure technical efficiency represents the failure of the firm to extract the maximum output from it adopt input levels, and hence it may be thought of as measuring the unproductive use of resources".

– that is related to choosing input mix given input prices. Furthermore, overall efficiency score for domestic banks was found to be around 83% which is the product of allocative efficiency of 92% and technical efficiency of 90%.

On the other hand, foreign banks were shown to be significantly less efficient with an overall score of 62% that breaks down into 82% for allocative efficiency and 71% for technical efficiency (Sathye, 2001, p 625).

The authors draw some policy implications from these results in that further financial deregulation to facilitate mergers between Australian banks is required. They argue that the implications of efficiency improvements are substantial: for a 10% increase in overall banking efficiency, the Australian economy would be able to save costs in excess of $4bn per annum.

Weill (2003) investigates banking efficiency in two countries Czech Republic and Poland in the aim of comparing the efficiency of 47 foreign-owned and domestically-owned banks. The study uses the Stochastic Frontier Analysis approach to compute cost efficiency scores. It follows Mester (1996) in including financial capital to accommodate for risk preferences. Methodologically, the analysis follows the two-step approach: firstly, efficiency estimates are found, and secondly, efficiency scores are regressed against a set of variables to investigate the sources of inefficiency. In the specification of the cost model, the study accounts for capital and country (dummy) variables only. Accordingly, the research of Weill (2003) suffers the same gap that most of past literature and that is ignoring risk variables. As for explaining the variation of efficiency results, the study regresses efficiency scores against a set of variables including: ownership, the ratio of loans to investment assets, the share of deposits in the total balance sheet, and the scale of operations as proxied by the size of bank assets. The study finds that foreign-owned banks are more efficient in both countries which was attributed to the type of ownership. The other factors: scale of operations and the structure of activities were found to have little significance in explaining the variation of efficiency estimates. The latter result may be questionable due to the fact that the study ignores accounting for risk comparatively, a significant gap in the literature that this study is attempting to bridge.

Drake and Hall (2003) explore technical efficiencies using non-parametric frontier estimation technique, DEA, in 149 Japanese banks using a cross-sectional data for the financial year ending in 1997. Findings indicate that controlling for risk – represented by the ratio of problem loans – has a significant impact on results. Drake and Hall (2003) conclude that the inclusion of risk seems to have a significant contribution in obtaining considerably different results from those of previous research as in Fukuyama (1993). That is, after controlling for problem loans mean technical efficiency rose from 72% to 86%. This indicates that X-efficiencies can be underestimated if risk is ignored. These findings are in line with Altunbas et al (2000) who find that efficiency estimates are quite sensitive to quality and risk factors in Japanese banking. Finally, on the size-efficiency relationship, technical efficiency was found to increase with bank size.

Drake et al (2006) examine technical efficiency of Hong Kong banking system by incorporating the effects of environmental and market factors –such as the Asian Crisis of 1997/98, and the regulatory reforms and increasing competition. To this end, a two-stage non-parametric estimation approach is utilized. The efficient frontier is therefore estimated using DEA followed by a second stag Tobit regression which allows for determining the effects of external factors on the estimated efficiency scores. The study obtains input and output data from Bankscope for the period 1995 – 2001 yielding 413 bank observations.

Drake et al (2006, p 1451) confirm the vitality of incorporating risk factors into the cost function as “failure to adequately account for risk can have a significant impact on relative efficiency scores”. In this context, risk is accounted for in terms of including loan loss provisions and equity capital. Overall, the study finds that Hong Kong banking system was affected by external factors with banks of different sizes and sectors being differently affected. Specifically, although the change in financial regulatory environment and the Asian crisis of 97/98 were found to have little impact on banks’ efficiency, the development in the macroeconomic cycle (declining GDP due to World economy slowdown in 2000) and the deterioration of house prices had a profound impact on banks’ efficiency, with smallest banks’ efficiency being least affected.

Moreover, the study reports a strong size-efficiency relationship so that after adjusting for external factors, smaller banks seemed to be more efficient with a score of around 92% compared to an average score of 88% for larger banks (Drake et al, 2006, p 1462). This is in contrast to previous US findings in that larger banks are found to be more efficient than smaller peers. Drake et al (2004) conclude with an important observation in that: failure to account for the impact of macroeconomic factors yields misleading efficiency scores over time for banks with different sizes and sectors.

Allen and Liu (2007) investigated cost efficiency of the six largest Canadian banks using a 20 years panel set from 1983 – 2003. The study specifies a Translog cost function. Average cost inefficiency was found to be around 10%, with largest banks enjoying the highest level of cost efficiency compared to relatively large banks. Allen and Liu (2007) attribute this result to management skills, higher loan quality, and most importantly, higher productivity. Moreover, a clear evidence of technological advances over time was found – technological progress was detected using linear and nonlinear (squared) time dummies. Banks seem to have taken advantage of technological advances to cut their costs hence improve their cost efficiency. This may have led to narrowing the dispersion of cost inefficiency between Canadian banks over time.

Ariff and Can (2008) investigate cost and profit efficiencies in Chinese banks using data on 28 banks with 230 observations over 1995 – 2004 and applying DEA. In line with earlier findings, the study finds profit inefficiency levels being significantly greater than cost inefficiency levels. A second-stage Tobit regression is employed to investigate potential drivers of inefficiency. Overall cost efficiency estimate was found around 0.8 suggesting that a typical bank wastes about 20% of its costs in producing a given output level relative to a frontier bank. On the other hand, overall profit efficiency level is found to average around 0.5 implying that an average bank seems to earn half of the profits that a frontier bank can make under similar conditions.

Furthermore, Chinese banks were stratified into 3 main asset groups: small (with assets less than $10 bn.), medium (with assets more than $10 bn. and less than $30 bn.), and large (with assets more than $30 bn.). In terms of size-efficiency relationship, the study reveals a positive and significant relationship for medium size banks, whereas the relationship is negative but not significant for small and large banks. This suggests, according to the authors, that medium-sized banks are more cost and profit efficient than small banks, however large banks are found to carry the lowest profit and cost efficiencies relatively.

Ariff and Can (2008) account for asset quality using the ratio of loan loss provisions/total loans, and credit risk by the ratio of loans/total assets. Both ratios were found to negatively and significantly correlate with efficiency, suggesting that banks with better asset quality (lower provisions to loans ratio) and lower credit risk (less loans in the asset portfolio) tend to operate with greater cost and profit efficiencies. However, their study found negative, but not significant, correlation between bank capital ratio (equity/assets) and efficiency. Non-interest income was also found to correlate significantly and positively with profit efficiency, suggesting that Chinese banks tend to make more profits (hence operate more profit efficiently) if they earn a higher proportion of fee and commission income. Cost-to-income ratio was found to significantly and negatively correlate with cost efficiency, confirming the expected result in that banks with greater control over cost which equally seek to enhance their revenues (i.e. lowering cost-to-income ratio as a result) can operate more cost efficiently.

Fitzpatrick and McQuinn (2008) examine profit efficiency for 55 large commercial banks (total assets > 1$ billion) operating in 4 different countries: Canada, the UK, Ireland and Australia over the period between 1996 and 2002. The study estimates the alternative profit function as specified by Berger and Mester (1997). Consolidated accounting data was drawn from Bankscope. The study focuses on large commercial banks in particular so as to avoid biasness in profit efficiency estimates due to differences in production technologies or other effects from being non-commercial banks. Banks were classified into 3 asset groups: small, medium, and large banks. The alternative profit function and the time-flexible technical inefficiency model suggested by Battese and Coelli (1995) were simultaneously estimated in a single-stage estimation approach, where inefficiency effects (determinants) are specified with the frontier functional form and estimated in one stage. However, it is also worth noting that the frontier function in their study follows the Translog specification.

Findings indicate that overall profit inefficiency for all sample banks operating in the four countries is at 31%, suggesting that the average bank in these four countries wastes about one-third of its attainable profits due to inefficiency. It should be noted though that the study finds evidence of the ‘bad luck’ influence besides the influence of ‘bad management’. Hence these inefficiencies should not be fully attributed to managerial deficiencies in achieving the maximum profits possible, although the ‘bad management’ factors were found to dominate.

In terms of size-efficiency relationship, large banks were found to have the lowest average profit inefficiency score of 0.2 (most profit efficient), while medium size banks were found to be the most profit inefficient (least profit efficient) at 0.39 of profits. Finally, small banks were found with a profit inefficiency score of 0.33. County-wise, banks operating in Canada seem to be less profit efficient than those operating in Australia. Furthermore, UK banks seem to be less profit efficient than their Australian counterparts possibly due to the fact that the UK has the highest average price of labour in the sample (the study reports that the UK has, on average, 58%, 20%, and 2% higher labour prices than those of Ireland, Australia, and Canada). Furthermore, within the UK banking sector, medium size banks seem to be the least profit efficient, while for non-UK banking systems, small size banks were found as least profit efficient. In general, the authors find that profit efficiency in the UK is, on average, lower than that of banks in Australia, Canada, or Ireland.

Berger et al (2008) investigate profit and cost efficiencies of the ‘Big Four’ state-owned Chinese banks using data on 38 Chinese banks (including the Big Four, foreign and other domestic banks) over 1994 – 2003 yielding 266 observations. The data covers 95% of the commercial banking assets in China (Berger et al, 2008, p 16). The Big Four constitute for nearly three-fourths of Chinese banking industry’s assets. Specifically, they controlled 72% of the total market share in 2003. A Translog functional form with time effects (i.e. including time dummies) is applied and estimated according to the two-stage approach. The study is motivated by recent reforms which the Chinese banking system has undergone in terms of partially privatizing these banks (IPOs) and allowing foreign investment to take minority ownerships in these banks. Therefore, the study aims is at examining the potential effects of such transformation on the Big Four's operational efficiency and compares it with foreign banks' efficiency.

Results show that mean profit and cost efficiency scores for the entire sample are 0.476 and 0.897 respectively suggesting more profit than cost inefficiencies. This is in line with earlier US-based findings (Berger and Humphrey, 1997). As for found the state-owned Big Four banks, the study found that they are the least cost efficient (scoring 0.892) and the least profit efficient (scoring 0.234 only) in the Chinese banking system.

The study attributes this considerable profit inefficiency to the high level of non-performing loans and weak revenues. On the contrary, the majority of foreign banks seem to rank among the most profit and cost efficient. Foreign ownership was found to significantly improve the efficiency of the Big Four over time. This is suggestive of the fact that ownership reforms seem to be strongly related with improving efficiency in Chinese banking system. The authors expect that these efficiency gains will be further extended by virtue of continued economic growth for the Chinese economy, and the increased flexibility of banking environment which does not only mean greater provision of credit, but also better and more efficient allocation of credit. Improved efficiency of the banking system in China seems to be another positive driver for economic growth for China, the study concludes.

Vu and Turnell (2010) investigate cost efficiencies of Vietnamese banking. The advantage of this study over past research is that it applies the monotonicity and concavity constraints in the estimation of the cost frontier. The study finds cost efficacies of these banks to be around 87%. The study uses the SFA approach in producing efficiency scores which has the advantage of reducing specification errors. The study subjects the cost functional form to many regulatory conditions such as linear homogeneity, monotonicity and curvature. The study develops the functional form such that it meets all regularity conditions at all sample data points. The latter step has shown that there are significant differences in the elasticities produced by the developed and pre-developed functional forms. The paper confirms that ignoring the incorporation of regulatory conditions produces misleading and biased efficiency results. To this end, this research adapts the practice of checking the functional forms that will be specified for these regulatory conditions. However, the paper misses on the comprehensive incorporation of the risk variables which this research has the advantage of. It is therefore worth investigating how the incorporation of the market and credit risk variables, having checked for the regulatory conditions, would impact the profit and cost efficiency scores. It is therefore believed that such contribution would have an added value to the efficiency literature.

Lensink and Meesters (2012) investigated cost efficiency for around 8000 banks using panel data in 136 countries over 10 years. Efficiency scores were estimated using the SFA approach. The functional form specification of this study accounts for macroeconomics variables including real interest rates, GDP per capital and GDP growth as well as other risk variables. On the risk side, the study incorporates Equity over Total Assets measure and the Return on Assets. As it seems, the study ignores accounting for credit and market risks comprehensively as this research comprehensively does. The paper also misses on checking for the very important regulatory factors that many past papers have shown the key impact of which on the accuracy of efficiency estimates. The other disadvantage of this paper is that it ignores estimating profit inefficiencies for the dataset used. This research estimates profit as well as cost efficiencies.

Das and Kumbhakar (2012) analysed the banking efficiency of Indian banks between 1996 and 2005 using the total productivity factor approach TFP. In terms of variable specification, the study considers factors including deposits, loans, number of accounts, classes of employees, and location of branches, however, it accounts for risk using capital adequacy ratio only. The functional form specified in this paper ignores accounting for risk comprehensively as well as ignoring testing the functional form for regulatory conditions. Such two aspects are fully considered and tested for in this research. This leads us to the conclusion that Das and Kumbhakar (2012) results might very much be questionable.

Shyu et al (2014) investigated the efficiency of the banking system in Taiwan, Hong Kong, and Mainland China using the 3-stage DEA model to estimate inefficiencies. The study accounts for a set of environmental conditions and other typical factors used in efficiency studies. It found some significant impact of the environmental factors on efficiencies, these included economic growth rate, annual growth rate of consumer prices, economic freedom and total population. In terms of risk, the study only accounts for capital adequacy ratio. All in all, this paper misses out on the very important factors of risks which this research comprehensively accounts for. In addition, the study ignores investigating the regulatory conditions which many previous studies have found a very important determining factor of the accuracy of the efficiency scores once these conditions have been checked for. Therefore, this research has a significant advantage over Shyu (2014).

2.5 An Extended Summary

The summary below gives a snapshot on the efficiency studies reviewed above which are most relevant to this research in terms of methodology and main findings. The summary is classified in three main subsections to deal with the reviewed European, Non-European and the rest of the world efficiency studies.

2.5.1 European Studies

Altunbas et al (2001) apply SFA to estimate a Fourier Flexible cost function describing a sample of EU-15 banks over 1989 – 1997 and find cost inefficiencies between 20 – 25%. Technical progress was also found to contribute in reducing cost by about 3%.

Cavallo and Rossi (2001) apply the two-stage estimation technique and specify a Translog cost function and the time-inflexible inefficiency model, Battese and Coelli (1992), to estimate cost efficiency for a sample of 442 banks in 6 major EU countries over 1992 – 1997. Overall cost inefficiency was found at an average of 15.64%. Findings also suggest that smaller banks are more cost efficient than larger banks. As for country-level efficiencies, banking system in Germany and the UK were found to be the most and least cost efficient respectively.

Vennet (2002) estimates Translog cost and profit functions over 1995 – 1996 using a sample with 2375 observations on banks from 17 European countries. Financial capital was used to proxy for default risk but other types of risk were ignored. Average cost and profit inefficiencies were found at about 20% and 30% respectively. Little evidence was found on the size-efficiency relationship (except for smallest banks which were found as the most profit efficient).

Maudos et al (2002) apply the two-stage approach to estimate Translog cost and profit functions for large banks in 10 EU countries over 1993 – 1996 using the DFA. Findings suggest more profit than cost inefficiencies. Country-wise, Belgium ranks as the least profit efficient, whereas banking systems of Luxembourg and Portugal are the most profit efficient. The UK banking system ranks as the 5th most profit and cost efficient in the EU-10, Austria ranks as the 1st, while Finland ranks as the 10th (last). Findings also show a non-linear relationship between bank size and profit efficiency, as only medium size banks (up to $10bn) were found with a negative and significant correlation with profit efficiency. In terms of risk (proxied by the standard deviation of ROA), evidence suggests positive (negative) correlation with profit (cost) efficiency.

Girardone et al (2004) estimate a Fourier Flexible cost function for Italian banks over 1993 – 1996. Overall cost inefficiency levels were found at about 13 – 15%, with declining inefficiency over time that is attributed to the impact of the SMP. No significant correlation between size and cost efficiency was found. Findings indicate that cost efficient banks seem to have higher interest income (in terms of net interest revenue or margin to total assets ratio), better control staff expenses (but no advantage in funding costs), more securities, more equity capital, and better asset quality. In terms of efficiency correlates, cost inefficiency was found to negatively correlate with capital strength but positively correlate with the level of none performing loans (NPLs).

Weill (2004) investigates cost efficiency of European banks over 1992 – 1998 using different parametric approaches. Under SFA, the study found no clear correlation between size and efficiency. Further, the study found significant and negative correlation between cost efficiency and cost-to-income ratio, that is, greater cost efficiency is associated with lower value of the cost-to-income ratio and vice-a-versa.

Casu & Girardone (2004) investigate profit and cost efficiencies of large banks operating in the EU-5 countries using a sample of around 2363 observations over 1993 – 1997. Average cost and profit inefficiencies were found at about 14% and 13% respectively.

Casu, Girardone and Molyneux (2004) estimate a Translog cost function for large banks in EU-5 countries using a panel of 2000 observations over 1994 – 2000. Results mainly indicate a general decline in total costs for the largest Italian and Spanish banks due to technological change. Results for Germany, France and UK were inconclusive. Bos and Kolari (2005) estimate a Translog profit and cost functions for large European and US banks using SFA over 1995 – 1999. Results show that European banks seem to have lower and more dispersed cost and profit efficiencies than US banks. Evidence from the US and Europe suggests that small banks are less cost and profit efficient than large banks. For European banks, average cost and profit efficiencies were found at about 0.95 and 0.721 respectively. For US banks, cost and profit efficiencies were higher at 0.976 and 0.749 respectively. This suggests more profit than cost inefficiencies in both contexts.

Kasman and Yildirim (2006) use SFA in a single-stage estimation approach to estimate Fourier Flexible cost and profit functions (truncated at the 2nd order terms) for commercial banks in 8 new member states of the EU in 2004. Average cost and profit inefficiencies were about 0.20 and 0.36 respectively. Cost efficiency was found to correlate positively with capital ratio, GDP growth, and the level of financial development (M2/GDP), whilst negatively correlated with deposit market concentration (HERF index) and inflation rate. On the other hand, profit efficiency was found to correlate positively with market concentration and the level of financial development but negatively with inflation and GDP growth. On the size-efficiency relationship, the study found no significant correlation between bank size and profit or cost efficiencies.

Altunbas et al (2007) use the two-stage estimation approach and apply SFA to estimate a Translog cost function to investigate the relationship between capital, risk and efficiency in European banks operating in 15 European countries over 1992 – 2000. The study specifies OBS as a third output in addition to loans and securities. Findings indicate that cost efficiency is negatively related to risk and capital ratio, but there seems to be mixed evidence on the relationship between liquidity and cost efficiency. As for country-specific indicators, results show a positive relationship between banking system liquidity, net lending (net loans to total assets ratio) and cost inefficiency. Lastly, cost inefficiency was found to negatively correlate with system-specific cost-to-income ratio.

Barros et al (2007) use the two-stage approach to estimate a Fourier Flexible cost and profit functions for a sample of European commercial banks (1384 observations) over 1993 – 2001. A logit regression is then applied to examine efficiency correlates. Results indicate that it is less likely for a bank to be worst performer (least cost and profit efficient) if it operates in Denmark, Germany, Portugal or Sweden. However, this probability increases if the bank operates in Finland, Luxembourg or the Netherlands. It is more likely for a bank to become among the best performers if it operates in competitive market, the study concludes. In terms of size-efficiency relationship, results indicate that smaller, loan-intensive banks are unlikely to be among the worst performers compared to bigger and more diversified banks.

Yildirim et al (2007) estimate Translog cost and 'alternative' profit functions for banks in 12 transition economies in Europe between 1993 and 2000 using SFA and the two-stage approach. Consistent with earlier evidence, they found more profit than cost inefficiencies. Average cost and profit efficiencies were found at 77% and 68% respectively. In terms of efficiency correlates, higher efficiency levels seemed to be associated with well-capitalized banks. The degree of competition was found to negatively impact profit efficiency whilst positively impact cost efficiency.

2.5.2 International Studies

2.5.2.1 US Studies

Berger et al (1993a) investigate X-efficiency of US commercial banks over 1984 –

1989 using the DFA estimation approach. X-inefficiencies were found to account for 20% of banking costs. Risk (proxied by standard deviation of ROA) seems to negatively correlate with profit efficiency. Strong and positive correlation is found between profit efficiency and size.

Spong et al (1995) use the two-stage approach to estimate a Translog cost model which incorporates proxies for credit, insolvency risk, and output quality. Results indicate that higher cost efficiency is associated with taking on more credit risk and holding higher levels of capital.

Clark (1996) estimates a Translog cost function for US banks over 1988 – 1991. Findings indicate that inefficiency estimates tend to be overestimated if risk (proxied by the standard deviation of ROA) is not accounted for (estimates dropped from 9% to 3% as a result).

Berger and Mester (1997) estimate Fourier Flexible cost and ‘alternative’ profit functions using a sample of about 6000 US commercial banks over 1990 – 1995 using SFA and the two-stage estimation approach. Average cost and profit inefficiencies were found at about 14% and 50% respectively, with tighter dispersion for cost efficiency estimates. No significantly positive correlation between profit and cost efficiencies was found. As for the size-effect, the study found that larger banks are more cost efficient but less profit efficient compared to smaller banks. Profit efficiency is negatively correlated with the ratio of purchased funds to total assets and the level of market concentration, whilst cost efficiency is negatively correlated to the latter factor. As for risk, the study found that riskier banks (i.e. those with higher SD-ROA or more volatile returns) tend to be less profit and cost efficient, an evidence of ‘bad management’.

Despite including 25 efficiency correlates in the second stage estimation, the model was able to explain about 7% of cost efficiency scores’ variance and about 35% of profit efficiency scores only.

Berger & De Young (1997) estimate a Fourier Flexible cost model for US banking using SFA over 1989 – 1994. The frontier function was estimated year-wise. Average cost inefficiency was found at about 8%. Cost inefficiency estimates were reproduced under a Translog specification and were found to significantly increase as much as twice the estimates under the Fourier Flexible. At the bank level, findings seem to favor the skimping hypothesis, suggesting that increases in cost inefficiency are associated with decreases in loan quality. The study also finds support for the moral hazard hypothesis as suggested by the negative relationship between capital strength and the level of nonperforming loans. At the industry level, evidence suggests that the bad management hypothesis seems to dominate the skimping hypothesis.

Clark and Siems (2002) apply the SFA and the two-stage estimation approach to estimate Translog cost and profit functions to examine the impact of accounting for off-balance sheet activities (OBS) on efficiency estimates using a sample of US banks over 1992 – 1997. Mean cost and profit inefficiencies were found at 14% and 35% respectively with OBS items specified. Cost efficiency estimates tend to be underestimated if OBS items are not specified. Results suggest that neither the relative size of off- to on-balance-sheet activities nor the bank size has any statistically significant correlation with either cost or profit efficiencies. The correlation of OBS activities with cost efficiency seems to be much stronger compared with profit efficiency.

Akhigbe and McNulty (2003) use SFA and the two-stage to approach estimate a Fourier Flexible profit function for US banks over 1990 – 1996. An overall negative size-profit efficiency correlation is found. Average profit efficiency for small and large banks was around 0.75 and 0.72 respectively. Profit efficiency seems to positively correlate with market concentration. Small banks seem to operate in concentrated markets, have higher loan quality, and have better access to retail deposits which gives them a funding cost advantage over large banks since large deposits are more expensive than retail deposits, besides, demand deposits are a source of fee-based income.

Akhigbe and McNulty (2005) also use SFA and the two-stage approach to estimate year-wise Fourier Flexible alternative profit function over the period of 1995 – 2001. A positive size-efficiency relationship is found. Profit efficiency is negatively correlated with capital ratio for medium (above $100 million and less than $1 billion) and large banks (assets above $1 billion), and with the level of nonperforming loan ratio. Profit efficiency, however, is positively correlated with deposit market concentration. Small US banks can significantly enhance their profit efficiency by maintaining high asset quality (low default rates), depend more on fee income, and operate in a concentrated deposit market. On the other hand, large banks can significantly enhance their profit efficiency by primarily using more leverage (depend less on equity).

2.5.2.2 Rest of the World Studies

Altunbas et al (2000) estimate a Fourier Flexible cost function using SFA approach for 139 Japanese banks over 1993 – 1996. OBS items are accounted for. Cost inefficiency was found to range between 5% and 7% and that these results are quite sensitive to risk and output quality factors.

Allen and Liu (2007) estimate a Translog cost function for largest Canadian banks over 1983 – 2003. Average cost inefficiency was found at about 10%. Technological advances were found to have a significant impact in reducing costs over time.

Fitzpatrick and McQuinn (2008) estimate a Translog ‘alternative’ profit function for large commercial banks in Canada, the UK, Ireland and Australia over 1996 – 2002. They apply SFA and the single-stage estimation approaches were used and the time-flexible technical inefficiency model suggested by Battese and Coelli (1995) was specified too. Average profit inefficiency was found at 31%. Evidence on the impact of ‘bad luck’ and ‘bad management’ was found with the latter’s effect dominating the former. A positive size-profit efficiency relationship was found. The UK banking system was found amongst the least profit efficient which the study attributes to its relatively higher labour costs.

Berger et al (2008) estimate Translog cost and profit functions for 38 Chinese banks over 1994 – 2003 using SFA and the two-stage approach. Average profit and cost efficiency scores are 0.476 and 0.897. The study found that state-owned banks are the least cost and profit efficient in the Chinese banking system.

2.6 Brief Summary

This brief summary provides the gist of the reviewed European and Non-European efficiency studies in a compact presentation.

European studies seem to mainly rely on 1990s data – except for Kasman and Yildirim (2006) who use 2004 data, estimate both Translog and Fourier Flexible functions, and predominantly use SFA and the two-stage approach. Cost and profit inefficiencies seem to range between 13 – 25% and 28 – 36% respectively. All European studies, to the best of the researcher's knowledge, seem to imply more profit than cost inefficiencies. There is some evidence of cost reduction due to technical progress (of about 3%). Many studies found no evidence on the size-efficiency relationship (in terms of profit or cost efficiencies), however, some studies found non-linear and negative relationships. In terms of risk, there is some evidence on positive (negative) correlation with profit (cost) efficiency.

As for efficiency correlates, evidence suggests that cost efficiency positively correlates with interest income, GDP growth, the level of financial development and asset quality, whilst being negatively correlated with the level of non-performing loans, deposit market concentration and inflation rate. With respect to the correlation with capital strength, there is evidence on a positive correlation with cost efficiency despite that few studies suggest a negative one. On the other hand, profit efficiency was found to correlate positively with market concentration and the level of financial development but negatively with inflation and GDP growth, nonetheless, the degree of competition was found to negatively impact profit efficiency.

Country-wise, banking systems in Germany and the UK were found to be the most and least cost efficient respectively. Other studies found that Belgium ranks as the least profit efficient, whereas banking systems of Luxembourg and Portugal are the most profit efficient. The UK banking system ranks as the 5th most profit and cost efficient in the EU-10, Austria ranks as the 1st, while Finland ranks as the 10th (last). Furthermore, some evidence suggests that it is less likely for a bank to be worst performer (least cost and profit efficient) if it operates in Denmark, Germany, Portugal or Sweden. However, this probability increases if the bank operates in Finland, Luxembourg or the Netherlands.

As for the **US evidence**, the vast majority of US studies seem to follow SFA and the two-stage approach to estimate a Translog or Fourier Flexible functions using 1990s data. Average cost inefficiency for US banks is found to range between 8 – 25% whilst profit inefficiency is found to range between 32 – 50%. Again, US evidence suggests more profit than cost inefficiencies. Cost efficiency estimates tend to be underestimated if OBS items are not specified. No significantly positive correlation between profit and cost efficiencies was found. There is mixed evidence on the size-profit efficiency correlation as some studies have found a strong and positive correlation while others have found a negative correlation. As for the functional form specification, cost inefficiency estimates under the Translog specification were found to significantly increase as much as twice the estimates under the Fourier Flexible specification.

Cost efficiency is found to negatively correlate with deposit market concentration, while profit efficiency seems to positively correlate with loan quality, the level of fee income, the size of retail deposits, and the degree of leverage. Profit efficiency is negatively correlated with the level of nonperforming loan ratio, capital ratio, and the ratio of purchased funds to total assets. There is mixed evidence on the correlation between profit efficiency and deposit market concentration.

In terms of correlation with risk, there is some evidence of a negative correlation with profit efficiency. More importantly, findings indicate that inefficiency estimates tend to be overestimated if risk is not accounted for. Also, riskier banks (i.e. those with higher SD-ROA or more volatile returns) tend to be both less profit and cost efficient, which is probably an evidence of the 'bad management' hypothesis.

Lastly, **rest of the world evidence**, mainly covering Australia, Canada, China, and Japan, seem to follow SFA and the two-stage approach to estimate a Translog or Fourier Flexible functions and mainly using 1990s and more recent data (up to 2003). Only Fitzpatrick and McQuinn (2008) seem to follow a very close methodology to that of this research by estimating an alternative profit function using SFA and the single-stage estimation approaches besides specifying the time-flexible technical inefficiency model suggested by Battese and Coelli (1995).

Cost inefficiency was found to range between 5% and 11%, whilst profit inefficiency was found to range between 13% and 53%. As in the US and European studies, international evidence confirms the fact that there is more profit than cost inefficiencies in banking. Cost efficiency estimates were also found sensitive to risk and output quality factors. Evidence on ‘bad luck’ and ‘bad management’ hypotheses was found with the latter’s effect dominating the former.

There is some evidence that technological advances have a significant impact in reducing costs over time. Generally, there seems to be a positive size-profit efficiency relationship. In terms of country-wise efficiencies, the UK banking system was found amongst the least profit efficient which the study attributes to its relatively higher labour costs. Also, state-owned banks are found to be the least cost and profit efficient in the Chinese banking system.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter explores the different methodological issues underpinning the empirical analysis. The chapter is split into two main sections.

With regard to the methodology related to efficiency analysis, the section starts by defining the main concepts and bank cost function. The discussion then follows on to examine the different frontier estimation techniques, with a particular focus on the stochastic frontier analysis approach (SFA) that is applied in this research. This is followed by specifying the modified profit and cost models to be estimated.

3.2 Efficiency Analysis

This section starts by introducing the concept of efficiency from a production frontier perspective followed by a reflection on the profit and cost frontiers and the related efficiency concepts (technical and allocative efficacies). The discussion then turns to examining bank production and cost functions in more details. The profit function is then specified from two perspectives (standard and alternative), followed by a detailed discussion on modelling and estimating technical efficiency term. The specification of banking inputs and outputs is also dealt with. The focus then shifts towards issues related to the specification of the functional forms (Translog and Fourier) with a detailed reflection on the regulatory conditions of the cost function. Parametric and non-parametric frontier estimation techniques are subsequently discussed.

3.2.1. Production Frontier

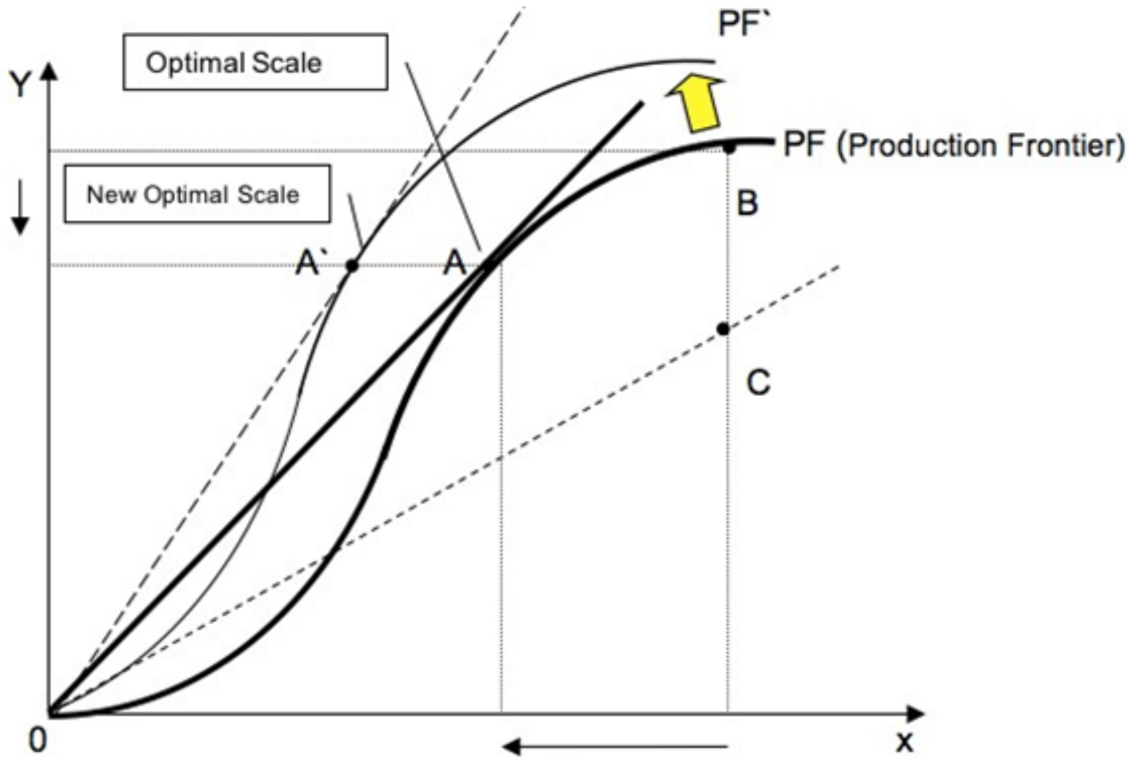
For a single-input and single-output production process, efficiency is simply calculated as the ratio of outputs to input levels given their prices. However, such simplicity cannot offer a comprehensive efficiency analysis in a multi-input multi-output setting. Thus, a mechanism of aggregating inputs and outputs to obtain efficiency indices needs to be applied which gave the origination of the frontier analysis framework (Coelli and Battese, 1998).

Frontier analysis is a benchmarking technique, which compares the performance of best-practicing banks in the sample, given inputs and outputs quantities and prices, according to a defined criterion – i.e. production, cost or profit concept – with the performance of the rest of the sample banks. The technique accordingly labels banks on the frontier as ‘efficient’, whereas observed banks that are off the frontier are labelled as ‘inefficient’. Thus, the efficient frontier contains all production subsets (i.e. feasible input-output combinations) of efficient banks.

It is essential in this context to differentiate between the concept of productivity and efficiency as they refer to different aspects. Productivity = outputs/inputs for one input one output production technology. But as banks operate with multi-input and -output production technology, measuring productivity is therefore referred to as total factor productivity, or shortly TP. Isolating production factors means considering the productivity of individual inputs such as labour or funds, in insolation yielding partial productivity measures (Coelli and Battese, 1998). Therefore, productivity reflects input/output relationship, whereas efficiency represents the relationship between total costs and input prices and outputs in the case of cost efficiency, and the relationship between profits and input prices and outputs for profit efficiency. However, these concepts are connected to one another as the cost function can be derived from the production function and vice-a-versa given the duality property of the cost function as we illustrate later.

It follows that banking production frontier reflects the maximum level of outputs obtainable given input levels. Therefore, a technically efficient bank would be positioned on the frontier, meaning that any observation lying below the production frontier is technically inefficient. Graphically, this is represented by a complicated multi- dimensional figure in which the frontier is depicted as a surface, however for simplicity reasons, this relationship is graphed in terms of one-input and one-output as shown in Figure 1 below.

Figure 1: Productivity, Technical Efficiency, and frontier shift over time (Coelli and Battese, 1998, p5)



Banks on the efficient production frontier exhibit, compared with other sample banks, the ability to produce the maximum possible amount of outputs given inputs, or produce a given level of outputs using the least possible inputs under similar conditions. As it can be seen from the production frontier above, producing at point C is inefficient as it entails lower level of output (y) for the same level of input utilized (x) compared to the efficient point B on the frontier. The slope of the production frontier (y/x) at a given point, represented by the tangent, yields the productivity achieved at that point. Hence any efficient observation beyond the optimal scale at point A will have a marginally decreasing productivity as inputs will be increasing more proportionately than outputs.

Point A on the production frontier corresponds to the trough or minimum stationary point on the corresponding long-run average cost curve LRAC (that is where marginal cost curve crosses LRAC – not shown here). It also corresponds to the maximum point on the average profit curve AP (where marginal profit curve crosses AP). Therefore, point A corresponds to the optimal scale (minimum average cost and maximum average profit). Production at an efficient point B is technically efficient because it lies on the frontier, but suffers scale diseconomies because it will be corresponding to a higher LRAC compared to that of point A. Production point C will have the same characteristics of B but with less output being produced for the same level of inputs used, hence it is technically inefficient and is also associated with scale diseconomies. Producing at any scale below A on the frontier means that the bank implies that the bank will be incurring higher average costs. Therefore, producing at point A is technically efficient.

Productivity could improve or decline over time which is reflected in upward or downward frontier shift respectively (Coelli and Battese, 1998). Figure 1 demonstrates an improved productivity causing the production frontier to shift upwards. At point A`, it becomes possible to produce the same level of (y) with less inputs used due to higher productivity (the slope of the frontier's tangent at A` becomes relatively greater). Point A` is technically efficient and corresponds to the new optimal scale of production.

Factors contributing to an upward shift in the production frontier mainly include technical development over time which entitles the bank to utilize inputs more efficiently, hence producing similar level of outputs but with lower level of inputs used. This can materialize, for instance, by exploiting its IT infrastructure to provide insurance products along with retail banking products, or by investing in new technologies that can reduce average costs in the long run as in inducing customers to utilize internet banking instead of branch-banking for instance. As can be seen, the production frontier represents an input-output relationship. On the other hand, profit and cost frontiers depict different relationship. The cost frontier represents relationship between costs and outputs, input prices, and other factors; while the profit frontier represents the relationship between profits and outputs, input prices, and other factors. Cost frontier banks are distinguished from other banks in the sample by their ability to produce the given level and mix of outputs under given input prices and other factors but with relatively lower costs. Banks off the cost frontier are technically less efficient because of the less effective use of their inputs or resources. Likewise profit frontier banks are distinguished for their ability to make maximum profits from a given level and mix of outputs under given input prices and other factors. Banks off the profit frontier are technically inefficient because of their excess costs or deficient revenues (Berger and Humphrey, 1997).

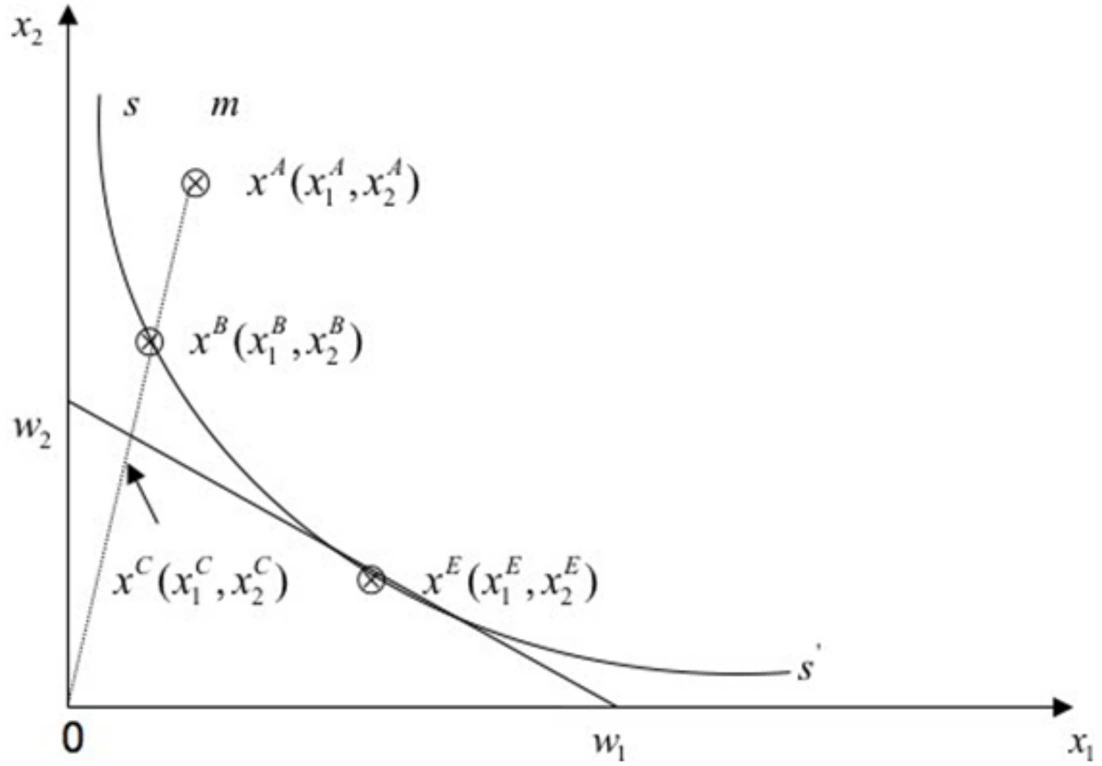
3.2.2 Technical and Allocative Efficiencies

Bank efficiency comprises of two main elements: technical and allocative efficiencies. Farrell (1957) explains that technical efficiency stands for the bank's ability to achieve maximum output by utilizing a given set of inputs at given input prices, or the ability to produce the observed output at minimal cost given input prices. On the other hand, allocative efficiency represents the bank's ability to use (allocate) inputs in optimal proportions under their given prevailing prices to produce a given level of output. The total effect of both efficiencies amounts to the so-called economic (overall) efficiency or X-inefficiency which was first introduced by Leibenstein (1966). According to Kumbhakar and Lovell (2000): X-inefficiency= technical inefficiency + allocative inefficiency.

Production technical efficiency measures the bank's ability to maximise outputs given input levels, reflecting the effectiveness in avoiding waste by producing maximum level of outputs given a set of inputs. From a profit maximization perspective, technical efficiency indicates the bank's ability to maximize revenues given output levels, input mixes, and inputs prices; hence advocating more effective utilization of inputs to increase profits. For a cost-minimizing bank, technical efficiency refers to the bank's ability to produce a given level of outputs using the minimal level of inputs given their prevailing prices.

Allocative efficiency, on the other hand, reflects the bank's ability to combine inputs and outputs optimally given their prices. In other words, allocative efficiency involves selecting the most cost-efficient inputs combination given their prevailing prices to produce a given level and mix of outputs.

Figure 2: Technical and Allocative Efficiencies



The figure above illustrates the concepts of technical efficiency and allocative efficiency (Fried et al, 1993, p91): The curve ss' represents the efficient isoquant specifying the production of a given output level Q using different inputs mixes**.** Any input mix situated on the efficient isoquant delivers similar level of output Q in a technically and allocatively efficient manner. The line w_1w_2 is the isocost (budget constraint) which represents the different compensations of inputs x_1, x_2 given their prevailing prices. Point x^A is the actual (observed) choice of inputs by a given bank. The technically efficient point that corresponds to the observed inputs choice, x^A, is the isoquant point x^B. x^B is a benchmark point employed to segregate total cost efficiency into technical and allocative efficiencies. The most technically and allocatively efficient point to produce Q given the observed input combination prices is x^E.

The cost of producing at point x^C is exactly the same to that of x^E as both lie on the same isocost w_1w_2, however, the input mix at point x^C is allocatively sub-optimal to x^E as it is not possible to obtain the output level Q that x^E can deliver for the same cost at the prevailing input prices (Fried et al, 1993). Allocative efficiency (AE) can therefore be measured by the ratio of inputs mix x^C to the optimal mix on the isoquant x^B such that: $AE_i = 0x^C / 0x^B$. This is referred to in the literature as Farrell measure of technical inefficiency (Farrell, 1957). AE is bounded between 0 and 1.

Technical efficiency (TE), on the other hand, can also be measured as: $TE_i = 0x^B / 0x^A$, suggesting that producing at x^A is technically inefficient given input mixes and prices. Input choices x^A and x^B correspond to similar input mixes as the relative proportion of x_2 to x_1 – which stands for the slope of the line $0m$ crossing the two choices – is the same for x^A and x^B. However, in absolute terms, x^A utilizes more inputs than the efficient choice x^B does to produce the given output quantity Q. This implies that the observed inputs choice at x^A is sub-optimal to that of x^B, which results in that producing at x^A is technically inefficient compared to producing at of x^B (Fried et al, 1993). TE is bounded between 0 and 1. Consequently, total efficiency or economic efficiency effect EE_i is the product of the two efficiencies such that (Coelli and Battese, 1998): $EE_i = AE_i + TE_i = (0x^C / 0x^B) \times (0x^A / 0x^A) \Rightarrow EE_i = 0x^C / 0x^A$, Where EE_i is bounded between 0 and 1.

It is worth noting that this research conducts profit and cost efficiency estimations under the assumption that all deviations from the frontier are attributed to technical inefficiency and the random error component following Coelli (1996). This is because allocative efficiency is difficult to disentangle from overall X-efficiencies since information on input mixes (combinations) is not available, therefore technical inefficiencies are assumed to account for all X-inefficiencies. That is why in the literature X-efficiency and technical efficiency are used interchangeably (Altunbas et al 2000)[8].

[8] The application of Shephard's Lemma to check for cost function monotonicity (by deriving factor share equations) suppresses allocative inefficiencies by default as it assumes that input levels and prices can change and affect

3.2.3 Bank Production and Cost Functions

The production function is a microeconomic concept representing the relationship between the bank's outputs and inputs. Banks transform inputs – labour, real capital and fixed assets – using their production technologies to generate outputs (different on- and off-balance sheet assets). These inputs can be utilized and transformed to produce the bank's outputs, and the effectiveness of doing so is conditional on the degree of allocative and technical efficiencies achieved (Shepherd, 1985).

Production technology can be expressed – in its basic form – by a setting of a single output and two inputs. This is illustrated by the Cobb-Douglas functional form which models output as a function of labour L and capital K (Besanko and Braetigam, 2005):

$$Q = f(L, K) \tag{1}$$

Early banking efficiency literature predominantly employed the Cobb-Douglas specification which treated banks as a single-output producing firm, but Cobb-Douglas handles a multi-product firm by estimating a separate cost function for each output (Mester, 1987). In this context, Cobb-Douglas clearly fails to establish a more realistic banking cost function that is capable of simultaneously accommodating for a vector of outputs produced.

Berger and Humphrey (1994) criticized the Cobb-Douglas specification since that it lacks the capability to allow for a U-shaped average cost curve as it can only estimate a decreasing, constant, or increasing average cost curve. Hence it is not suitable for detecting the existence of for a set of banks that are heterogeneous in scale. Another drawback is that Cobb-Douglas assumes unit elasticity of substitution between inputs, ignoring the fact that some inputs may have elastic or inelastic rate of substitution given their relative prices. They further highlight a serious weakness of this model in that it denies the all-important property of duality of the cost function and the

costs, holding input mix as constant (Berger and Mester, 1997).

production function[9].

Advances in the production economics literature have led to the departure from Cobb-Douglas model and the development of the Constant Elasticity of Substitution production function CES (Kim, 1985). The latter overcomes the shortfalls of Cobb-Douglas in that it permits for any degree of substitution between inputs rather than restricting it to unit elasticity and provides more general specification for modelling firms' production functions. Nonetheless, it only accommodates a production function with one output which makes the CES significantly constraining when using it to model a multi-output setting as is the case for banking production technology.

Molyneux (1996) stresses this by indicating that the CES offers a highly restrictive functional model and that it does not impose a homogenous elasticity of substitution between each input pair. Moreover, the CES, as does the Cobb-Douglas functional form, suffers from an additional limitation in that it does not allow for estimating a U-shaped average cost curve – a property that hinders the reliability of analysis (Gilbert, 1984). Such serious drawbacks of Cobb-Douglas and CES functional forms have prompted further research in the area of exploring more suitable and flexible functional forms.

This has materialized in the work of Christensen et al (1973) in developing the so-called Translog cost function. The Translog function is a flexible functional form hence it is superior to Cobb-Douglas and CES models as it overcomes their restrictiveness related to limiting the elasticities of substitution of the various factors of production (inputs) (Kim, 1985, p 1). According to Mitchell and Onvural (1996, p 178) "Translog represents a second-order Taylor series approximation of an arbitrary function at a point".

The Translog cost function is obtained by applying Taylor series expansion in input prices and output quantities around a specific point of a log-linear quadratic

[9] To analyse efficiency, the object of interest is the production function, i.e. how inputs are combined to produce outputs. However, the data required for its direct estimation is generally not available; moreover, there are methodological issues related to how to deal with differences in product quality. Therefore, cost and profit functions are alternatively estimated that basically contain all the relevant information involved in the production function, i.e. they are duals of the production function (Amel et al, 2004, p 8).

approximation of a multi-product arbitrary 'unknown' cost function that is defined as: $\ln TC = f(\ln y_i, \ln w_i, t)$. Translog is a linear combination of input and output terms, but is superior to Cobb-Douglas in that it allows for estimating U-shaped cost function as it contains squared output terms – hence a trough is mathematically possible (Humphrey, 1990). It is worth noting that Cobb-Douglas allows for a U-shaped cost function as it could involve quadratic terms, however the Translog is more flexible as it does not only accommodate for quadratic terms, but it also contains second-order interactive input prices and output quantities terms which permits Translog to be more flexible than Cobb-Douglas (Humphrey, 1987 and 1990). It also accommodates for multi-output multi-input production technology where each input and output is treated as a separate variable, and permits for any degree of substitution elasticities between inputs. Most importantly, Translog specification allows for marginal costs (i.e. derivative of cost function w.r.t. outputs) to vary with output levels (Humphrey, 1990).

For any production function, there is a unique cost function that can be derived from it using the economic objective of cost-minimization. There are two properties that are related to the cost functions in this respect: Shepherd's Lemma and Duality. The Shepherd's Lemma[10] enables the derivation of input demand functions from the cost function. Cost function duality, on the other hand, is a microeconomic property that establishes the link between production and cost functions, hence making the cost function dual to the production function. Duality is important because modelling the production function in banking is difficult, therefore the cost function is modelled instead and certain restrictions are imposed on it to make it dual to the production function so inference can be made about the production technology of banks.

Diewert and Kopp (1982) illustrate that any cost function can be derived with no prior knowledge of the underlying production function. As data on input prices and output quantities is normally more readily available than that on input quantities, banking production functions are therefore difficult to estimate. For this reason, the cost function is constructed and analysed instead (Diewert and Kopp, 1982). Christensen

[10] Shepherds' Lemma is a microeconomics property which utilizes the concept of cost minimization to establish the link between the cost function and input demand functions. That is, input demand functions can be generated by taking the rate of change of (or the derivative of) the cost function with respect to input prices ($\partial TC/\partial w_i$) according to Mankeiw and Taylor (2006).

and Green (1976) explain the rationale for this in that the estimation of the production function is preferred when output quantity is determined endogenously, however, estimating the cost function is far more attractive if output quantity is determined exogenously as it is mostly the case for banking.

Thus, total costs are defined as the sum of inputs expenditures and modelled as a function of output quantities and all input prices. This can be illustrated by the following relation (Schotter, 2008):

$$TC = w_1.L + w_2.K \tag{2}$$

Where w_1, w_2 are prices of labour L and capital K. Also, total costs can be expressed in terms of the product of average total costs (ATC) and the quantity produced:

$$\begin{aligned} TC &= ATC \times Q \\ TC &= \frac{TC}{Q} \times Q \\ TC &= \frac{w_1.L + w_2.K}{Q} \times Q \end{aligned} \tag{3}$$

This leads to a unique relationship between the cost and production functions as follows:

$$TC = \frac{w_1.L + w_2.K}{Q} \times f(L, K) \tag{4}$$

This formula indicates that total costs are a function of input quantities (L, K), input prices (w_1, w_2) and output quantity $Q = f(L, K)$.

Production theory describes the methodology by which an arbitrary production function for banking institutions can be approximated, yet it does not precisely specify an algebraic functional form for the bank cost function. Rather, it states certain requirements that a well-behaved cost function should satisfy.

A ‘proper’ cost function has to conform to five main regulatory conditions as suggested by production theory. These are: positive cost elasticities (outputs’ marginal costs), monotonicity in input prices, concavity in input prices, linear homogeneity in input prices, and symmetry in input prices and output levels. These requirements, according to Shepherd (1970), Caves et al (1980), Gilbert (1984), Kim (1985), Jorgenson (1986), Diewert and Wales (1987), Mester (1987) and Le Compte and Smith (1990), involve the following:

(1) The cost function needs to be nonnegative (i.e. positive marginal costs of outputs for each observation), monotonic or non-decreasing in inputs (i.e. total costs should be monotonically be factor shares: labour, capital, and funds shares), non-positive own-price elasticities[11] to satisfy concavity in factor (input) prices for different output levels, and linearly homogenous in input prices to achieve duality to the production function,

(2) The cost function should comprise a reasonable number of unknown parameters to be estimated, and

(3) The cost function has to be flexible such that it provides approximation to an unknown arbitrary function. To achieve this, the cost function should be characterized to allow for a second order local approximation parameter (so that it can be differentiated twice to explore possible stationary points.

These conditions will carefully be tested and accommodated for in the specification of the preferred cost function prior to conducting the estimation and producing efficiency and results.

[11] A necessary and sufficient condition to test for concavity is to check if the symmetric matrix of the second-order factor price term to be negative semi-definite as demonstrated by Diewert and Wales (1987). This is thoroughly checked for in the both empirical chapters.

3.2.4 Model Specification

In this section, the general form of production function is specified then the cost and the alternative profit functions are discussed. This is followed by a discussion on the specification of the inefficiency term and the associated necessary tests. To begin with, the production function is specified according to Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). This model implies regressing output quantity against input levels with an error term that comprises two components: a random error component and a technical inefficiency component. This model is shown as follows:

$$\ln y_i = x_i\beta + (v_i - u_i) \quad , \; i = 1, 2\ldots N \tag{5}$$

Where:

y_i is the output for the i^{th} bank.

x_i is a vector of input quantities for the i^{th} bank.

β is a set of coefficients to be estimated.

v_i is the random variable assumed to be $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ and is independent of the inefficiency component u_i.

u_i is the nonnegative random component that is assumed to ascribe technical inefficiency where $u_i \overset{iid}{\sim} N^+(0, \sigma_\mu^2)$.

This model can be extended to accommodate for different distributional assumptions of u_i, a panel data set, and for cost and profit function specifications (Coelli, 1996). The discussion now turns to focus on the cost function and the alternative profit functions which will be estimated in the corresponding empirical chapters.

3.2.5 The Cost Function

Berger and Mester (1997) indicate that banking efficiency analysis is best approached through three main economic concepts, these are: cost, standard profit, and alternative profit efficiencies. Their argument stems from the concept of 'economic optimization' as it accommodates for the state of the bank's production technology as well as market's conditions (competition and prices). According to Berger and Humphrey (1997), Berger and Mester (1997), and Mester (1996), the cost function is specified to derive cost efficiency which measures the ability to produce the observed output level at minimum cost given input prices.

Berger and Mester (1997, p 898) illustrate that the cost function implies specifying total costs as a function of: variable outputs, variable input prices, fixed netputs[12] (inputs or outputs), environmental factors, random error, and a given level of inefficiency. This is expressed as:

$$C = C(w, y, z, v, u_c, \varepsilon_c) \tag{6}$$

Where:

C are the variable costs.

w is a vector of variable input prices.

y is a vector of variable output quantities.

z is the quantity of any fixed netputs.

v is a set of environmental or market variables that can affect performance.

u_c denotes the inefficiency term that could cause the bank-specific costs to rise above the efficient frontier.

ε_c represents the composite random error that may temporarily increase or decrease bank-specific costs. It is assumed to incorporate the effect of variables' measurement error, good or bad luck...etc.

---

[12] Netputs are basically inputs or outputs which are taken to be fixed and/or where the prices of inputs are difficult to measure, such as equity capital as in Berger and Mester (1997). Netputs in this research include: Off-Balance Sheet items (the credit equivalent of), Physical Capital, and Equity Capital (Berger and Mester, 1997, p 915).

The composite random error term is assumed to consist of two components: a random component v_c and the inefficiency component u_c. The inefficiency term u_c and the random term v_c are assumed to have different distributional assumptions so they can be separated. The two components therefore are multiplicatively separable form the rest of the independent variables in the cost function.

The idea behind separating the composite error term is to attribute the portion of the dependent variable' variation (be it cost or profit) that could not be explained by the model's independent variables (output quantities, input prices, macroeconomic variables…etc) to a random error component capturing the effects of luck, measurement error…. etc., and another one-sided (positive) component capturing inefficiency effects. So, in the case of the cost function, a bank's cost inefficiency is attributed to using more input quantities (not specified in the cost function) to produce a given level of outputs under given input prices and other variables (which are specified in the cost function).

Thus, once the cost function is specified, the functional form of the frontier in equation (7) can be specified assuming the two components of the composite error term. Accordingly, taking the natural logarithms of both sides of equation (6) yields the frontier's functional form as follows:

$$\ln C = f(w, y, z, v) + \ln u_c + \ln \varepsilon_c$$

(7)

Subsequently, each bank's cost efficiency indicator $CostEFF^b$ can be defined as the ratio of the minimum predicted costs in the sample to the actual predicted costs of the specific bank. This is defined as (Berger and Mester, 1997, p 898):

$$CostEFF^b = \frac{\hat{C}^{min}}{\hat{C}^b} = \frac{\exp[\hat{f}(w^b, y^b, z^b, v^b)]\exp[\ln \hat{u}_C^{min}]}{\exp[\hat{f}(w^b, y^b, z^b, v^b)]\exp[\ln \hat{u}_C^b]}$$

$$= \frac{\hat{u}_C^{min}}{\hat{u}_C^b}$$

(8)

Cost efficiency score denoted as $CostEFF^b$ represents bank b's cost of producing the output vector (y) with no cost inefficiency ($u = 1$ or $\ln u = 0$) under similar conditions in terms of: input prices(w), environmental variables (v) and fixed netputs (z). $CostEFF^b$ ranges between 1 and infinity.

Put simply, a bank-specific cost efficiency score $CostEFF^b$ provides the percentage of costs which the observed bank (b) can reduce by using its resources more effectively, relative to the best-practicing bank in the sample operating under similar conditions, to produce its given level of outputs. For instance, a $CostEFF^b$ of 110% suggests an excess in the costs or waste in the resources utilized of 10% relative to the best-practicing bank on the cost frontier in the sample.

3.2.6 Standard and Alternative Profit Functions

Berger and Mester (1997) illustrate that the concept of profit efficiency examines how close a bank is to the maximum Possible profit achieved by the best-practicing bank in the sample, given certain conditions (input prices, output prices, environmental variables…etc). There are two competing concepts in this respect: the standard profit and the alternative profit functions.

The standard profit function assumes perfectly competitive output markets hence banks are price takers rather than price makers in both output and input markets. On the other hand, the alternative profit function assumes that banks can have some power in determining output prices. That said, the standard profit function specifies profits as a function of input prices and *output prices*, while alternative profit function specifies profits as a function of input prices and *output quantities*. Therefore, the specification of the standard profit efficiency function considers input and output prices as exogenous (outside the control of banks and defined by the market), whereas output quantities are assumed to be endogenous (can be influenced by banks).

Berger and Mester (1997) point out that it is necessary to analyse profit efficiencies alongside with cost efficiencies. This is because the cost function specifies output levels in the model as given, or as a control (independent) variable, and therefore having no impact on the inefficiency term in the cost function. This is despite that these output quantities may be at inefficient levels in the first place. Accordingly, the cost function allows for capturing *inefficiencies in the choice of inputs only*. On the other hand, as the standard profit function specifies output prices as a control variable, it therefore leaves output level and mix as well as inputs level and mix to influence revenues and therefore influence the inefficiency term in the profit function. This effectively allows for capturing *inefficiencies in the choice of outputs and inputs* in response to output prices and other control variables specified in the cost function. However, if output quantities are specified in the profit function, yielding the alternative profit function as will be shown, this allows for capturing inefficiencies in the choice of output prices in response to changes in output quantities and the other control variables in the profit function.

With regard to the specification of the standard profit function, Berger and Mester (1997, p 899) provide the following form in natural logarithm terms:

$$\ln(\pi + \theta) = f(w, p, z, v) + \ln(u_{\pi}) - \ln(\varepsilon_{\pi}) \tag{9}$$

Where:

π[13] stands for the variable profits of the bank.

[13] Variable profits incorporate all interest, fee, commission income earned on the variable outputs minus variable

θ is an added constant so the logarithm is taken of a positive number.

p is the vector of variable output prices.

$\ln(\varepsilon_{\pi})$ is the random error.

$\ln(u_{\pi})$ stands for inefficiency which reduces profits.

Standard profit efficiency index is then calculated as the ratio of a bank b's predicted[14] actual profits to the predicted maximum profits of the best-practicing bank in the sample. Hence:

$$\text{Std } \pi \text{ EFF}^{b} = \frac{\{\exp[\hat{f}(w^{b}, p^{b}, z^{b}, v^{b})] \times \exp[\ln \hat{u}_{\pi}^{b}]\} - \theta}{\{\exp[\hat{f}(w^{b}, p^{b}, z^{b}, v^{b})] \times \exp[\ln \hat{u}_{\pi}^{\max}]\} - \theta}$$

$$= \frac{\hat{\pi}^{b}}{\hat{\pi}^{\max}}$$

(10)

As such, Std π^{b} can range between 0 and 1. So for instance if a bank's Std π^{b} was 80%, this implies it is losing 20% of the profits due to excessive costs, deficient revenues, or both.

The standard profit function has a disadvantage given the inherent assumption of the exogenous nature of the output prices vector (p) implying the absence of the bank's control over (p). Such assumption is relaxed by the concept of alternative profit function which permits for the assumption that banks can operate under imperfectly competitive market structure, hence banks can exercise control over output price vector (p). Accordingly, output quantities vector (y) are specified instead of output prices (p) in the profit function. Consequently, the profit function becomes defined as:

costs.

[14] These are predicted because the model's parameters describing the dependent variable (actual profits) are estimated or predicted by the model.

$$\pi = \pi(w, y, v, u)$$

(11)

The alternative profit function considers output levels (y) as exogenous or given, i.e. determined by market equilibrium between supply and demand, whereas output prices are accounted for as an endogenous variable that is determined by the bank. Therefore, under the alternative profit concept, the economic objective of maximizing profits (π) is achieved by controlling output quantities $(y)$and input prices(w). Profit inefficiency therefore arises from deficient revenues (output prices and mixes) or excessive costs (resulting from excessive use of input levels or deficient input mixes), or both. This is because both factors are not specified in the profit function but left to be captured by the inefficiency term in the function.

Following Berger and Mester (1997, p 899), the specification of the Alternative Profit functional form in natural logarithm terms implies that:

$$\ln(\pi + \theta) = f(w, y, z, v) + \ln(u_\pi) - \ln(v_\pi)$$

(12)

Where:

π[15] stands for the variable profits of the bank.

θ an added constant so the logarithm is taken of a positive number.

y is a vector of variable output quantities.

$\ln(v_\pi)$ is the random error component.

$\ln(u_\pi)$ stands for inefficiency which reduces profits.

It is probably worth noting that the cost and alternative profit functions have the same independent variables specifications, with different dependent variables. Also, v_π is subtracted from profit inefficiency u_π so that profit efficiency score becomes bounded between 0 and 1. Each of the bank-specific alternative profit efficiency score is

[15] Variable profits incorporate interest, fee, and commission income earned on the variable outputs minus variable costs.

calculated by taking the ratio of a given bank's predicted actual profits to the predicted maximum profits of the best practicing banks. This is given by the following (Berger and Mester, 1997, p 901):

$$\text{Alt }_{\pi}\text{EFF}^{b} = \frac{\{\exp[\hat{f}(w^{b}, y^{b}, z^{b}, v^{b})] \times \exp[\ln \hat{u}_{a\pi}^{b}]\} - \theta}{\{\exp[\hat{f}(w^{b}, y^{b}, z^{b}, v^{b})] \times \exp[\ln \hat{u}_{a\pi}^{\max}]\} - \theta}$$

$$= \frac{\hat{a\pi}^{b}}{\hat{a\pi}^{\max}} \qquad \textbf{(13)}$$

This research specifies the alternative profit function instead of the standard profit function to analyze technical efficiency. Opting for the alternative profit function specification stems from the arguments of Berger and Mester (1997).

The authors list a set of assumptions that when one or more hold, the alternative profit function should be specified instead of the standard profit function if:

I - **There are considerable unmeasured differences in the quality of bank products and services.** This is because the alternative profit function concept accommodates the additional revenues generated by higher quality output. Higher output quality is associated with higher costs which are normally associated with higher prices. The specification of the alternative profit function can accommodate for this since it specifies variable output quantities (y) as given but leaves output prices (p) to freely fluctuate reflecting the different levels of output quality and affect the level of profit inefficiency (since inefficiency is part of the model's residual). This is relevant especially when discrepancies in output quality amongst sample banks are evident. One aspect of output quality can be represented by the level of nonperforming loans for instance. As far as the sample of European banks used in this research, these differences in loans quality are evident as the ratio of loan loss provisions to total loans ranges between [0.27%, 6.36%] for the sub-sample of 541 banks that are the subject of profit efficiency estimation. This relatively wide dispersion in output quality underpins the application of alternative profit function –banks differ in loan quality by almost 23-fold as the range indicates.

II - **Scale of production is not completely variable.** This assumption accommodates for the fact, given the strong heterogeneity in banks' asset size, that it takes a considerable amount of time for smaller banks to grow and 'catch up' with mega-banks through growth and mergers and acquisitions. Therefore, it is significant to allow the profit function to level the playground for banks with substantial scale differences. The Standard profit function overlooks this issue as it does not control for output scale whereas alternative profit function specifies output levels as a factor determining profits, hence reducing scale bias in efficiency results when the sample has considerably heterogeneous bank sizes. Although this analysis is confined to big European banks only, sample scale heterogeneity is very evident. Bank sizes involved in the sample range over [985 - 1,125,500] million euros in assets. This further justifies the application the concept of alternative profit to analysis profit efficiency.

III - **Banks can exercise a degree of market power implying imperfect competition.** The concept of standard profit function ignores that banks may not operate in a perfectly competitive market as it specifies output prices as given, i.e. banks can sell variable levels of output without changing prices, whereas in fact some banks can exert some market power hence affecting prevailing market prices. Such perfect completion assumption is economically unsound as banks that are producing below the efficient scale cannot – under the standard profit function – lower their prices to increase output and exploit cost savings, therefore their standard profit efficiencies can be underestimated. This is overcome by the concept of alternative profit function as prices are free to vary. Berger and Mester (1997) find strong evidence of banks exerting market power and suggest that banks with larger market shares have some influence over prices, and therefore they can lower rates they pay to depositors (Berger and Hannan 1989, and Hannan, 1991).

The data used in this research shows strong evidence of imperfectly competitive markets for many of the European countries involved. This is demonstrated using Herfindahl index for loans which refers to the size of loan market share controlled by certain number of banks in a given market. For example, the average Herfindahl index over 2008 - 2018 for Denmark was 47% which represented the market share of 5 big commercial banks out of 26 commercial banks operating in Denmark over that period in total. Also, for Switzerland, 2 big commercial banks out of 41 controlled 56% of the loan market over the specified period. However, the UK market is found as competitive where 34 big commercial banks out of 96 controlled only 12% of the loan market. In light of this, the application of alternative profit function instead of standard profit function is necessary so as to accommodate for the considerable degree of market concentration in a number of European banking markets.

IV - **Output prices are measured inaccurately.** Standard profit function relies on obtaining accurate output prices to be specified in the function so as not to produce biased profit efficiencies (Pastor and Serrano, 2005). Berger and Mester (1997) argue that banking output prices are very likely to be inaccurately measured which gives the specification of profit function according to alternative profit function an edge over the standard profit function, not to mention the fact that it is quite difficult to obtain these prices. Because the alternative profit function addresses the difficulty in measuring output prices, it allows for output quantities (rather than output prices) to explain the variation in bank profitability (Akhigbe and McNulty, 2003).

Further, raw data on banking output prices has weak explanatory power in relation to profits. In fact, using US data Humphrey and Pulley (1997), modelled both output prices and quantities to explain the variations of the dependent variable, profits, and tested the null hypothesis that all output price and quantity parameters are zero. The null for output prices variable was not rejected, while that of output quantities was strongly rejected.

Alternative profit function assumes that output prices are endogenous to the bank’s profits (chosen by the bank) and therefore should not be modelled as an independent variable to explain variations in profits. On the other hand, output quantities are assumed to be exogenous to profits since they are determined by market equilibrium, therefore output quantities are specified in the profit equation as a control variable. Accordingly, banks are assumed to operate under imperfect output markets and therefore can exercise price differentiation. The alternative profit concept is believed to be more consistent with the nature of output markets that large banks operate in where exercising some market power is possible (Bos and Kolari, 2005).

3.2.7 Estimating Technical Inefficiency

Efficiency is measured as the deviation from the frontier where best-practice banks operate. The frontier is established by estimating the cost or profit function using data on the observed banks. Inefficiencies are subsequently estimated as deviations from the frontier after isolating the random error effect. Evaluating the performance of banks as such requires using two different optimization concepts: cost minimization and profit maximization (Berger and Mester, 1997). A bank is labelled inefficient if its profits (costs) are lower (higher) than the best-practice bank after removing the effect of the random error (Vennet, 2002).

Under the alternative profit model, profit inefficiency arises from deficiencies in revenues or excess costs given output quantities and input prices. It follows that technical inefficiency (u_{it}) is bounded between 0 and 1 because all inefficient banks operate below the efficient profit frontier. Profit efficiency simply measures how close a bank is to generating the maximum obtainable profits given the prevailing input prices, output levels, and other factors. A profit efficiency score of 0.9 means that the bank is effectively earning 90% of best-practice profits, that is, the bank is losing 10% of possible attainable profits due to excessive costs, deficient revenues, or both (Vennet, 2002).

Under the cost model, cost inefficiency arises from sub-optimal choices of input quantities given input prices and output quantities. It follows that the inefficiency term is bounded between 1 and infinity since all inefficient banks operate above the efficient cost frontier (Bos and Kolari, 2005). A cost efficiency score of 1.10 for instance means that the bank is wasting 10% of its resources to produce a given level of outputs relative to a best-practice (frontier) bank.

The profit and cost models are estimated using SFA (stochastic frontier analysis) approach which basically applies the MLE (maximum likelihood estimation) to produce the required estimates. The stochastic frontier approach SFA has been widely applied in the literature as in Allen and Rai (1996), Mester (1996), Berger and Mester (1997), and Altunbas et al (2000). SFA labels a bank as cost/profit inefficient if its observed costs/profits are higher/lower than the predicted costs/profits for an efficient (frontier) bank producing comparable input-output combination and that the difference cannot be explained by statistical noise (Altunbas et al, 2000).

To produce bank-specific efficiency scores, the cost/profit efficient frontier is obtained by estimating the corresponding cost/profit function according to SFA. The cost/profit function will have a composite error term that comprises of a two-sided random component and a one-sided inefficiency component.

The reason why inefficiency effects are estimated conditional on the error term's distribution is that the former is unobserved, while the latter is observed. As shown in equation (14), data point technical efficiency is defined as $TE_i = \exp(-u_i)$. Given that the vector of parameters in the frontier function are estimated using MLE, the only effect that can be observed is that of the composite error term ε_i, yet the bank-specific inefficiency effect u_i is unobserved. That is why u_i is predicted conditional on the value of ε_i.

Accordingly, Jondrow et al (1982) calculates observation-specific technical inefficiency using the distribution of the inefficiency term conditional on the estimated distribution of the composite error term by the following relationship:

$$E[u_i \mid \varepsilon_i] = -\gamma\varepsilon_i + \sigma_A\{\frac{\phi(\gamma\varepsilon_i / \sigma_A)}{1-\Phi(\gamma\varepsilon_i / \sigma_A)}\} \tag{14}$$

Where:

$\sigma_A = \sqrt{\gamma(1-\gamma)\sigma^2}$,

$\varepsilon_i = \ln(y_i) - x_i\beta$, and

$\phi(.)$ is the density function of a standard normal random variable.

By substituting the estimated parameters in (14), technical inefficiency can then be evaluated. $E[u_i \mid \varepsilon_i]$ is alternatively expressed as $E[\exp(-u_i) \mid \varepsilon_i]$, as in Battese and Coelli (1998). This is because the frontier functional form is expressed in logs so obtaining estimates for the inefficiency term requires exponentiating (de-logging) it. This implies that estimating $\exp(u_i)$, rather than u_i, conditional on ε_i. In this context, bank-specific technical efficiency TE is then specified according to Battese and Coelli (1998)[16] as:

$$TE_i = E[\exp(-u_i) \mid \varepsilon_i] = \frac{1-\Phi(\sigma_A + \gamma\varepsilon_i / \sigma_A)}{1-\Phi(\gamma\varepsilon_i / \sigma_A)} \exp(\gamma\varepsilon_i + \sigma_A^2 / 2) \tag{15}$$

Once the log-likelihood function (that is unique to the frontier specified) is estimated, the distribution of the composite error term can be identified, and estimates of the model's parameters can be used in equation (15) to predict banks-specific technical efficiency estimates. Bank-specific cost efficiency can therefore be produced by dividing the estimated technical efficiency of the given bank by the predicted technical cost efficiency of the frontier bank. Similarly, bank-specific profit efficiency can be produced by dividing the estimated technical efficiency of the frontier bank by the predicted profit technical efficiency of a specific bank.

The literature postulates a number of distributional assumptions for the technical inefficiency term including: truncated-normal, half-normal, exponential, and gamma. The technical inefficiency model suggested by Battese and Coelli (1995) assumes a truncated-normal distribution for the inefficiency term. Nevertheless, there has been a considerable body of efficiency studies that assume the half-normal, rather than the truncated normal, distribution for u_i (e.g. Altunbas et al 2000, Vennet 2002, and Girardone et al 2004).

[16] FRONTIER and STATA apply the technical efficiency predictor defined in equation **Error! Reference source n ot found.** by replacing MLE estimates for the unknown parameters.

Vennet (2002) for instance uses the half-normal assumption $u_i \overset{iid}{\sim} N+(0,\sigma_\mu^2)$ after testing both the half-normal and exponential distributions. The author found no statistically significant difference between the two sets of efficiency scores under the different distributional assumptions. Furthermore, the study by Altunbas and Molyneux (1996) tested the half-normal, truncated normal, exponential and gamma distributions and found that efficiency estimates derived via SFA methodology are, to a large degree, insensitive to the assumption of the inefficiency term's distributional specification.

Ritter and Simar (1997) support employing the relatively simple one-parameter densities, namely, half-normal and exponential distributions, as opposed to more flexible distributors, such as gamma and truncated normal. Their argument for this relies on the existence of a relatively strong correlation between mean efficiency scores under different distributional choice, suggesting that mean efficiencies are found to be relatively insensitive to u_i distributional specification. The correlation is even stronger between mean efficiencies under half-normal and exponential distributional assumptions, leading to the conclusion that the choice between the latter two distributions is principally irrelevant. It is also established that the case of half-normal distribution for technical efficiency is most frequently used in empirical research as Coelli et al (1998) state.

This research tests for truncated-normal, exponential and gamma distributions for the inefficiency term and found that none of these specifications would enable the stochastic cost or profit function to converge under the MLE technique even after 16000 iterations. The models converged and the corresponding log-likelihood functions displayed concavity (i.e. they were possible to evaluate) under the half-normal specification only. Therefore, and given the above-mentioned reasons, the half-normal distribution for the inefficiency term is specified.

3.2.8 Modelling Inefficiency Effects

The literature applies two prominent models describing the inefficiency term: Battese and Coelli (1992) and Battese and Coelli (1995). The two models basically differ in the time-variances assumption of the inefficiency term, and the estimation procedure according to which each model is evaluated (two- or single-stage procedures).

Under Battese and Coelli (1992) model, the stochastic production frontier requires a panel set and assumes that bank-specific inefficiency effects u_{it} as non-negative random variables which vary systematically over time (i.e. u_{it} have a specific time trend that is imposed by the model), and follow half-normal (truncated at zero) distribution. The model is estimated following the two-stage procedure whereby the efficient frontier is first estimated, and then the predicted bank-specific inefficiencies (with a specific time-trend for the predicted bank inefficiency effects being imposed) are regressed against a set of explanatory variables in a second stage regression.

Battese and Coelli (1992) model is expressed as follows (Coelli, 1996):

$$\ln y_{it} = x_{it}\beta + (u_{it} - v_{it}) \qquad , i=1, 2\ldots N; \ t=1, 2\ldots T \tag{16}$$

Where

- y_{it} is the output for the i^{th} bank in the t^{th} period.
- x_{it} is a vector of input quantities for the i^{th} bank in the t^{th} period.
- β is a set of coefficients to be estimated.
- v_{it} are random variables assumed to be $v_{it} \overset{i.i.d.}{\sim} N(0,\sigma_v^2)$ and independent of the inefficiency component u_i.
- u_{it} are nonnegative random variables that are assumed to ascribe technical inefficiency where $u_i \overset{iid}{\sim} N^+(\mu,\sigma_\mu^2)$ truncated at zero such that:

$$u_{it} = [u_i \exp(-\eta(t-T))] \tag{17}$$

Where:

- T is the time variable representing the panel's last period, and t is a specific time period in the panel.
- η eta is an unknown parameter to be estimated which determines the direction of the inefficiency term's progression over time. If η>0 (<0), the exponential expression in equation (17) can be raised to positive (negative) power causing bank-specific inefficiencies to have an upward (downward) sloping time trend which indicates larger (smaller) inefficiencies as time progresses.

The maximum likelihood estimation procedure MLE applied to the model considers the parameterization of Battese and Corra (1977) where σ_u^2 and σ_v^2 are re-parameterized to be expressed in terms of the variance of the composite error term: $\sigma^2 = \sigma_u^2 + \sigma_v^2$, and a new parameter gamma $\gamma = \sigma_u^2 / \sigma^2$ which represents inefficiency variance as proportion of the composite error's variance, making γ ranges between [0, 1]. This parameterization is useful as the range of γ becomes the basis for a grid searched across the parameter space of γ to provide starting value for iterative maximization process in the MLE estimation (Coelli, 1996).

It is worth noting here that allocative inefficiency in u_{it} is suppressed (assumed away) following Coelli (1996) so that the interpretation of the inefficiency term would reflect the degree of technical inefficiency involved, otherwise understanding deviations from the frontier become blurred as both technical and allocative inefficiencies can account for deviations from the frontier.

Specifying the inefficiency term under Battese and Coelli (1992) model is discarded for the fact that it does not allow for the single-stage estimation procedure that is more consistent than the two-stage procedure, and because it forces the time progression of the inefficiency term to be systematic and therefore allows little time-flexibility in the specification of the inefficiency term. For these two reasons, this research applies the time-flexible Battese and Coelli (1995) inefficiency model (Kumbhakar and Lovell, 2000). Battese and Coelli (1995) model has similar specifications to that of the Battese and Celli (1992) defined in equation (17) except that the technical inefficiency effects

u_{it} are directly influenced by – or is simultaneously estimated with – the efficient frontier and modelled on a set of explanatory variables z_{it} that are believed to explain the variation of u_{it}.

Following Aigner et al (1977), the cost (or profit) frontier's composite error term is specified as: $\varepsilon_{it} = u_{it} + v_{it}$ (or $\varepsilon_{it} = u_{it} - v_{it}$) as discussed earlier, where the inefficiency component u_{it} and random error component v_{it} are assumed to be independently distributed. u_{it} is assumed to follow a one-sided positive distribution usually half-normal[17]: $u_{it} \sim N^{+}(0, \sigma_u^2)$ capturing the inefficiency effects in the model, and v_{it} is assumed to follow a two-sided normal distribution where $v_{it} \sim N(0, \sigma_v^2)$ capturing the effects of statistical noise in the data (Altunbas et al, 2000, p 1608).

Because the values of the inefficiency term u_{it} are unobservable whereas the values of ε_{it} are, u_{it} is therefore predicted conditional on ε_{it} estimates. This is achieved by applying the conditional expatiation of u_{it} conditional on the observed value of ε_{it} (Coelli, 1996). Accordingly, observation-specific predicted inefficiencies are obtained as $E(u_{it} \mid \varepsilon_{it})$. Once predicted inefficiencies u_{it} are obtained, bank-specific inefficiency scores (EFF_{it}) are produced as (Coelli, 1996, p8):

$$EFF_{it} = E(TC_{it}^{*} \mid u_{it}, x_{it}) / E(TC_{it}^{*} \mid u_{it} = 0, x_{it}) \qquad \textbf{(18)}$$

Equation (18) produces cost efficiency scores which can take values between 1 and infinity, where TC_{it}^{*} is the actual or observed total cost for a bank i at time t. The

[17] The literature postulates different distributional assumptions for the inefficiency term including half normal, truncated normal, exponential, and gamma. Altunbas and Molyneux (1994) tested the different distributions including the half normal, truncated normal, normal-exponential and gamma distributions and found that all distributions yield similar efficiency levels using German banking data. This research tests many of thesis different and finds that the estimated profit and cost models can only converge if half-normal distribution is assumed.

denominator represents the observed costs of the cost-efficient banks (frontier banks) featuring no deviation from the frontier due to technical inefficiency. x_{it} are the cost frontier variables. If the frontier equation is expressed in logs, then TC^{*}_{it} will effectively be equal to $\exp(TC_{it})$ and u_{it} will be equal to $\exp(u_{it})$.

On the other hand, producing profit efficiency scores follows similar methodology to that of equation (18) with profits P^{*}_{it} being specified instead of total costs TC^{*}_{it}. Profit efficiency scores will accordingly take the values up to unity.

Once the predicted inefficiency values u_{it} are produced, u_{it} can then be regressed against a set of explanatories (z_{it}) variables in order to understand the drivers of technical inefficiency. Furthermore, to allow for time-variances, this research specifies both linear and non-linear forms of the time dummy as z_{it} variables as shown in the equation below. Following Battese and Coelli (1995, p 327), the inefficiency term u_{it} is modelled against a set of bank-specific variables such that:

$$u_{it} = \alpha + \beta_i z_{it} + \omega t_t + \vartheta \tfrac{1}{2} t_t^2 \tag{19}$$

Where:

i is the number of banks in the panel.

t represents time periods in the panel.

z_{it} represents a set of explanatory variables introduced to explain the variation in u_{it} A full description of these bank-specific variables is presented in Table 3 (page 199).

$t_t, \tfrac{1}{2} t_t^2$ represent a linear and a non-linear form of time.

β_i , ω, ϑ are a set of coefficients to be estimated.

The time-flexible inefficiency model specified in (19) will be estimated simultaneously for the profit and cost frontier models in a single stage estimation procedure following

SFA methodology. Further details on this will be presented in each of the respective empirical chapters.

3.2.9 Technical Inefficiency Test

This test investigates whether the frontier model has inefficiency effects or not. Coelli (1998) suggests this test which involves examining whether the variance of the inefficiency term is significantly different from zero. Hence $H0:\sigma_u^2=0$ is tested against the alternative $H1:\sigma_u^2>0$ since inefficiencies can only be nonnegative. Findings in both empirical chapters indicate that the profit and cost frontiers specified have significant inefficiency effects.

3.2.10 Testing for Stochasticity

A stochasticity test is conducted to find out whether the frontier model should be considered as stochastic in the first place. Coelli (1996) suggests testing the model to be estimated for stochasticity prior to estimating the model as a stochastic frontier. This is empirically achieved by estimating the frontier only, excluding the inefficiency term's explanatory variables. In this research, the test is applied to the profit and cost frontiers. Stochasticity test involves testing the significance of the parameter $\gamma=\sigma_u^2/\sigma^2$ which implies testing whether the variance of the inefficiency term constitute a statistically significant proportion of the variation of the composite (or aggregate) error term.

Accordingly, the null states that the technical inefficiency is not stochastic, that is $H_0:\gamma=0$, which effectively suggests testing whether σ_u^2 is statistically different from zero in relation to σ^2. This is tested against $H_1:\gamma>0$ (as technical inefficiency is always assumed as being nonnegative) which implies that the frontier is stochastic. Not rejecting the null suggests that the frontier's parameters can consistently be estimated using ordinary least squares OLS where the inefficiency term can be removed from the model (Coelli, 1996, p 5). To empirically perform this according to Gutierrez et al (2001), the frontier model is estimated by maximizing the likelihood function and subsequently using the log-likelihood ratio test. Results in both empirical analyses show that the null hypothesis that $\gamma = 0$ is comfortably rejected at 1% critical level, suggesting that the frontiers are stochastic because the inefficiency term has a significant variation in relation to the variation of the composite error term hence the models should be estimated using SFA.

3.2.11 The Two-Stage and Single-Stage Frontier Estimation

The two-stage approach involves estimating the efficient frontier in the first stage, yielding the predicted efficiency scores, followed by a second stage regression where the predicted inefficiency is modelled against a set of possible exogenous determinants (Coelli, 1996, p 5). The two-stage approach carries an inherent contradiction: the first stage involves the specification and estimation of the frontier functional form and the prediction of the inefficiency effects. The estimation of the frontier in this first stage assumes that these inefficiency effects are identically distributed as one-sided error terms. Nonetheless, the second stage of this process involves regressing the predicted inefficiency effects on a set of potential determinants (hence assuming that inefficiencies come from a two-sided error distribution) which basically contradicts the assumption of an identically distributed one-sided error term assumed in the stochastic frontier's specification (Iyer et al, 2005 and Fitzpatrick and McQuinn, 2008).

On the other hand, the single-stage estimation procedure overcomes the inconsistency of the two-stage's distributional assumptions. It estimates the efficient frontier and regresses the explanatory variables against the inefficiency term simultaneously (Coelli, 1996), therefore maintaining the same distributional assumption for the inefficiency effects throughout the estimation process. Therefore, Battese and Coelli (1995) and Coelli (1996), conclude that the two-stage estimation procedure is likely to produce efficiency estimates that are less efficient to those of the single-stage. It is worth noting that testing for the preferred frontier specification under the single-stage against the two-stage procedures is impossible since the models are non-nested (Coelli 1996). Consequently, the choice of this research of applying the single-stage estimation approach is based on avoiding the contradicting assumptions of the two-stage approach which can impact the accuracy of the inefficiency effects' estimates.

3.2.12 Analysing Profit and Cost Efficiencies

As discussed earlier, *cost* is defined as a function of input prices, output levels, cost inefficiency term, and a random error. *Cost efficiency* is the ratio of a bank's true cost to the cost of best-practicing bank producing the same bundle of outputs. *Profit* is also defined as a function of input prices, output levels, profit inefficiency term, and a random error. *Profit efficiency* is the ratio of the bank's true profits to those of the best practicing bank producing the same bundle of outputs.

Therefore, the economic objective of profit maximization embodies the objective of cost minimization. This is because maximizing profits requires maximizing the vector (mix) of production levels – which maximizes revenues – and minimizing costs as well such that profits (revenues – costs) can be maximized. However, the argument underpinning the rationale for analysing both profit as well as cost efficiencies stems from the findings of Berger and Mester (1997) and Rogers (1998) in that there is no positive correlation between profit efficiency and cost efficiency. This suggests that a bank being profit efficient does not implicitly suggest that it is cost efficient as well, but on the contrary, the bank could be cost inefficient (i.e. incurring higher costs) despite it being profit efficient. Moreover, Berger et al (2008) indicate that it is important to estimate profit efficiency besides cost efficiency since the latter neglects operating revenues and loan losses, whereas profit efficiency clearly accommodates for these elements.

Furthermore, Berger and Mester (2003) also assert the significance of investigating both profit and cost efficiency concepts as this would provide a more comprehensive understanding of the different aspects affecting operational efficiency. Berger and Mester (2003, p 29 – 30) indicate that "banks tried to maximize profits by raising revenues and well as reducing costs. Over time, banks have offered wider varieties of financial services and provided additional convenience. These additional services or higher service quality … may have raised costs but also raised revenues by more than these costs increase".

This clearly implies that a bank being labelled cost inefficient does not necessarily mean that it is profit inefficient, as it may have managed to raise revenues more proportionately than costs to make up for the additional expenses involved in providing higher quality products and services for instance.

On the other hand, a cost-efficient bank may be profit inefficient since it is not maximizing its profits when there are unexploited opportunities to increase output in quality or quality, hence to generate greater revenues. Profit efficiency is enhanced by maximizing revenues and minimizing costs (operational costs, financial costs, and the level of loan loss reserves). Therefore, if costs are minimized but revenues are kept constant when it is possible to maximize them more proportionately than cost increases – as in the case of exploiting potential – this leads to lower profit efficiency as a result, accordingly, a cost-efficient bank is not necessarily a profit efficient one. The best-case scenario is to minimize costs and to maximize revenues.

Rogers (1998) explains that estimating both profit and cost efficiencies rather than considering profit efficiency only is useful because profit efficient banks may effectively be cost inefficient by offsetting higher expenses with higher revenues. This is possible by exploiting market power to impose higher prices, or by altering the mix of its output so as to generate higher revenues than competitors. This is probably further exposed when the risk-return paradigm is considered: some banks might take excessive risks to boost short-term profits at the expense of investing less in risk management process to cut costs. This is referred to as risk management 'skimping' according to Berger and De Young (1996). Consequently, it seems that to obtain the full picture, efficiency needs to be assessed from the two perspectives, and therefore examine the factors driving each in an attempt to investigate any possible correlation between profit and cost efficiencies as far as this sample banks are concerned.

3.2.13 Bank Inputs and Outputs Specification

3.2.13.1 Intermediation and Production approaches

The literature defines two prominent methodologies related to the specification of the banking institution's inputs and outputs: the intermediation approach and production approach (Berger and Humphrey, 1997). A brief introduction of both approaches is presented and followed by a stream of arguments for employing the intermediation approach in this research.

To start with, the production approach entails that a bank uses labour and capital to generate deposits, loans, and securities accounts. Therefore, deposits are treated as an output of the production process. Outputs are proxied by the number – rather than the currency value – of deposits and loans. This is because the production approach regards banking production as an account-servicing process that involves processing payment transactions, loans documentation, insurance applications etc. It follows that the application of such approach would require an in-house type of detailed service flow data, which is normally difficult to access. Therefore, this hurdle is overcome by using the number, rather than the value, of deposits and loans accounts, insurance policies and so forth as a proxy for the bank's outputs (Berger and Humphrey, 1997).

Studies adopting the production approach argue that the bulk of banks' operating costs are associated with the number – rather than the monetary value – of loans, securities, and deposit accounts processed (Benston, 1965 and Humphrey, 1985). However, Humphrey (1990) demonstrates that specifying costs as operating costs versus total costs would largely produce biased estimates. Technological advances have made it significantly less costly to process large number of accounts hence operating costs associated with these operations have been reduced significantly. Humphrey and Vale (2004, p 1691) have found, based on a panel of 120 Norwegian banks over 1987 – 1998, that average reduction in operating costs associated with the shift from paper-based to electronic payments has been around 40%. Thus, accounting for operating costs only can seriously underestimate the bank's costs.

On the other hand, the intermediation approach considers labour, capital, and deposits as inputs to produce loans, investments and securities (Sealy and Lindley, 1977). Accordingly, banks are viewed as intermediaries between fund-holders and fund-seekers, implying that total costs should accommodate not only operating costs, but also interest expenses paid on deposits. Earlier empirical studies tested the specification of bank output with and without deposits. They found that the sum of squared errors was lower by specifying deposits as inputs to bank production as opposed to dropping deposits as inputs (Gilligan and Smirlock, 1984).

Furthermore, deciding on which approach to measure bank outputs according to also hinges on the type of cost (operational or total cost) that is going to be modelled as a left-hand side variable in the cost function. In this context, the production approach focuses on using operational costs[18], while the intermediation approach uses total costs instead. The rationale for this different treatment of cost is that, since the production approach assumes deposits as outputs not inputs, deposit-related interest expenses are accordingly excluded from the aggregate pool of input expenses. Conversely, since the intermediation approach assumes deposits as an input to the bank production process, cost of input factors will have to accommodate for interest expenses borne on deposits hence total costs as modelled as a function of inputs process and output quantities.

Sealy and Lindley (1977) indicate that accounting for deposits as inputs is consistent with the economics of banking production. The production process, in economic terms, involves creating a product that has an added value to the value of original inputs utilized. Consequently, deposits can have an “added value” for the bank only if they undergo a maturity transformation that turns deposits into loans, investment and other types of outputs. In line with this logic, deposits can only be introduced as inputs to the bank’s production process. Drake and Simper (2002) follow the intermediation approach because they see that deposits are intermediated into different categories of earning assets using the inputs of capital and labour. Berger et al (1987), Noulas et al (1990), and Humphrey (1990) illustrate that applying the production approach would result in rather inflated estimates for small banks as they rely heavily on deposits as a source of funding, given their limited access wholesale funds. Consequently, this would put large banks at an inherent disadvantage as they have limited core deposit base compared to their smaller peers. What is more, employing the intermediation approach and considering bank’s total cost perspective is consistent with the fundamental economic objective of cost-minimization in that banks seek to minimize their total costs, rather than their operating costs only (Dietsch, 1993).

[18] Operating costs include all expense required to keep the bank operational such as labour wages and other admin expenses.

Equivalent to the Intermediation Approach is the Assets Approach methodology. Likewise, the Asset Approach treats banks as intermediaries of financial services between savers (liability holders) and borrowers (fund users), rather than just producers of deposits and loan accounts as in the production approach. Under the Assets Approach, the bank earning assets, such as loans and other investments, are classified as outputs, whereas labour, deposits and capital are considered as inputs (Favero and Papi, 1995). It should be noted that the assets and intermediation approach are generally referred to interchangeably in the literature.

Another approach suggested by Berger and Humphrey (1991) and Bauer et al (1993) is the Dual Approach which combines the production and introduction approaches in treating deposits. The Dual approach specifies interest paid on deposits as an input while considering the quantities of deposits as an output assuming that account services provided to customers can be approximated by the quantity of deposit accounts, rather than their monetary value.

The dual approach has very narrowly been applied in the literature. To the best of the researcher's knowledge, the only study that has applied the dual approach was by Cavallo and Rossi (2001). Considering deposits as outputs can produce biased results as explained earlier and found by Humphrey (1990). Furthermore, the argument against this approach is that deposits are used, especially by commercial banks, as an important source of funding to their assets and investments – probably the Northern Rock crisis is a chief example of how significant deposits can be as part of the bank's capital. Therefore, it is questionable whether deposits should be accounted for as outputs at all.

This crosses with the arguments of Shaffer and David (1991) and Jagtiani et al (1995) discussed earlier who demonstrate that deposits should be included as inputs, not outputs.

Efficiency literature recognizes two more approaches concerning the specification of commercial banking inputs and outputs, these are: The User Cost Approach 'UCA', the Value-Added Approach 'VAA'. The main difference between these approaches lies in specifying bank's inputs, however, both agree on classifying loans as outputs (Berger and Humphrey, 1997). A brief explanation followed by a reflection on the inappropriateness of each in relation to this research will be presented, thereby supporting the argument of using the intermediation approach.

In principle, the VAA and UCA are distinguished from the other approaches in that they are not related to the bank's macroeconomic functions (Favero and Papi, 1995). To be more specific, the UCA categorizes bank's assets and liabilities groups as inputs and outputs depending on their net contribution to the bank's revenues (Hancock 1991 and Ausina, 2002). For this, the UCA approach is impractical from an outside research perspective. The impracticality associated with the UCA rests on the fact that it requires relatively detailed income and cost data on the value of cash flows generated by or paid on each asset and liability category in order to classify these items as outputs and inputs respectively.

With regards to the Value-Added Approach 'VAA', balance sheet items are classified as inputs or outputs according to the 'importance' of the value each item adds. However, deciding whether a balance sheet item adds an 'important' or 'unimportant' value hence classifying it as an output or otherwise is a fairly subjective matter, which is probably the main weakness of the VAA (Favero and Papi, 1995).

Some studies argue (Favero and Papi, 1995, p 388) that the intermediation or assets approaches fall short of comprehensively accounting for most of the services provided by banks such as securities underwriting. Nevertheless, this research overcomes this by including the credit equivalent of off-balance sheet items (OBS), along with the traditional banking outputs (loans and other earning assets).

3.2.13.2 The importance of accounting for OBS items

This research recognizes the significance of accounting for OBS items as a third output in the cost and profit functions as ignoring them can underestimate banking outputs and bias estimates accordingly. OBS items are introduced in the frontier functions as credit equivalent accounts according to Basel I credit conversion criteria.

This way, OBS items are perceived to have the equivalent impact as the different classes of assets and liabilities on the balance sheet in terms of their revenue-generating or cost-bearing characteristics, therefore implying that OBS items have approximately the same origination, monitoring, and control costs as loans[19] (Berger and Mester, 1997, p 913).

Moreover, several European studies (Vennet, 2002, Maudos et al 2002, and Casu and Girardone, 2004) appear to ignore accounting for Off-Balance Sheet items OBS as an output[20]. It is argued that unless off-balance sheet activities are accounted for as an output, the bank's total output will significantly be underestimated especially in the case of large banks (Altunbas et al, 2001). Clark and Siems (2002) assert that "given the growth of OBS activities, estimating bank cost and profit efficiency without incorporating these activities may not be accurate or meaningful. Omitting OBS activities could seriously understate actual bank output and seriously bias empirical estimates of the relationships between bank size and both cost and profit efficiency" (p 988)[21]. For these reasons, OBS items are accounted for in this research as a third output. The recent Subprime financial crisis is clear evidence of the extent to which banks are relying on OBS activities to further diverse their earnings, including securitization and credit derivatives which effectively allows them to take debt off their balance sheets (FT, 2007a).

3.3 Functional Form Specification

This section is concerned with illustrating the concepts of the functional forms utilized in the two empirical chapters. Banking efficiency and literature predominantly applies

[19] It is recognized, as pointed out by Clark and Seims (2002), that accounting for the credit equivalent of OBS activities does not capture the volume of OBS activities or their ability to generate income or costs. Clark and Seims (2002) use an aggregate measure of OBS activates to this end. They apply the volume of net noninterest income as a proxy for the aggregate impact of OBS activities. Non-interest income is accounted for in this research under the 'other earning assets' output.

[20] Although Casu and Girardone (2004) argue that due to possible variations in cross-country accounting practices which could result in scale bias, efficiency estimations are very likely to be misleading for the considerable significance of income generated via OBS activities.

[21] Clark and Siems (2002, p 988) also explain that "between 1990 and 1999, commercial banks increased the notional value of their financial derivative activities by more than 600 percent. Noninterest income, which is heavily influenced by OBS activities, has increased as a percentage of total income from 19 percent in the late 1970s to nearly 46 percent by 1999. Thus, it is increasingly important to include these activities in any evaluation of the efficiency and competitive positioning of commercial banks" (p 988).

two main functional forms[22]: the Translog form [23](Translog) and the Fourier Flexible[24] form (Fourier).

3.3.1 The Translog Form

Briefly, the Translog function is based on a second-order Taylor expansion of an unknown true cost function around an unspecified data point (White, 1980). Translog is a generalization of the single-input single-output Cobb-Douglas production function resulting from Taylor series expansion. McAllister and McManus (1993, p 391) further explain that "the Translog cost function was originally developed as a local approximation to some unknown 'true' underlying cost function. The interest in the approximation was motivated by the fact that it does not impose restrictions on elasticities at the point of approximation". This is in comparison to the more restrictive Cobb-Douglas form. Translog and Cobb-Douglas can both estimate a U-shaped cost function given that they basically are quadratic functions, however, Translog is more flexible because of the introduction of the interactive inputs-outputs terms in addition to the second-order own-interactive input and output terms. The weakness of the Translog specification stems from the fact that it provides a poor global approximation when the data is spread away from the mean. This can cause misspecification error which can directly affect the accuracy of estimates. To elaborate, in spite of the flexibility of Translog, White (1980) explains that the least squares estimates of Translog do not generally correspond to Taylor series expansion of the underlying function at an expansion point, hence the least squares estimates are biased. This is because the second-order Taylor expansion of the unknown true cost function around an unspecified point produces an expansion around that point that is far away from the actual data upon which the least squares estimates are produced. This led White (1980) to the conclusion that estimating the Translog cost function generally produces

[22] There are other functional forms: Cobb-Douglas, Box-Cox, Quadratic, and Leontief. However, these are quite narrowly applied in the literature due to the superiority of the Translog, and the semi-parametric (Fourier) forms. Semi-parametric refers to the Translog form nested in a Fourier form (Fenn et al, 2008).

[23] Studies used Translog form to estimate include: Benston, Hanweck, and Humphrey 1982; Gilligan, Smirlock, and Marshall 1984; Gilligan and Smirlock 1984; Berger, Hanweck, and Humphrey 1987; Cebenoyan 1988; Hunter, Timme, and Yang 1990; Noulas, Ray, and Miller 1990, Kolari and Zardkoohi 1991, Berger and Humphrey 1991, Gropper 1991, Humphrey (1993), Berger et al (1993), Mckillop et al (1996), and Allen and Liu (2007).

[24] Studies applied the Fourier Flexible include: McAllister and McManus (1993), Mitchell and Onvural (1996), Berger et al (1997), Girardone, Molyneux, and Gardener (2004), Huang and Wang (2001), and Humphrey and Vale (2004).

biased estimates.

To confirm the biasness of Translog estimates, McAllister and McManus (1993) demonstrate that studies estimating bank efficiency using Translog cost functions usually find that economics of scale exist for relatively small banks, whereas larger banks are found to display diseconomies of scale. These results were found to be biased (i.e. the mean of estimator distribution is not the true parameter – that is, on average, the estimated values do not converge to the true values). McAllister and McManus (1993) find that big banks were found with constant returns to scale once Fourier terms are considered in the cost function. They conclude that Translog gives a poor approximation when applied to all bank sizes. Moreover, Berger and Humphrey (1997) also argue that Translog normally gives a poor approximation to banking data when the data is not close to the mean scale (widely spread), and that Translog imposes a symmetric U-shaped average cost curve in logs.

In line with this, Berndt and Khaled (1979), Jorgensen and Fraumeni (1981), and Jorgenson (1986) state that estimated Translog cost functions normally do not satisfy concavity in prices. This is confirmed by the findings of the first empirical chapter where the Translog function specified to estimate cost and alternative profit efficiencies failed the concavity test, which has led to the consideration of the 1st order Fourier terms to provide a better fit to the data so that the cost function eventually becomes concave. The concavity condition is verified and achieved for all functional forms utilized in both empirical analyses as will be shown later.

Consequently, Mitchell and Onvural (1996) observe that this has serious policy implications as the inadequacy of the Translog does not only cast doubt on the conclusions of previous studies that used this form, it also challenges policymakers who have been guided by these studies in making antitrust and bank merger decisions to revise their policy rules.

3.2.2 The Fourier Flexible Form

The Fourier Flexible form is an expansion of a periodic function containing an infinite sum of trigonometric terms of sine and cosine terms (Rudin, 1976 and Gallant, 1981). The Fourier form “is a globally flexible approximation since the respective sine and cosine terms attached to the Translog from are mutually orthogonal over the [0, 2π] interval” (Humphrey and Vale, 2004, p 1678). This implies transforming or rescaling input prices and output quantities such that they become expressed in trigonometric expressions that will span in [0, 2π] interval. The linear combination of the resulting sine and cosine terms has the mathematical property of being orthogonal (i.e. independent from one another). The flexibility of the Fourier stems from the orthogonality property of the sine and cosine terms, which implies that the sines and cosines are mutually independent or exogenous when specified in the Fourier function as they span in [0, 2π] space with no overlapping. Thus, because of the mutually orthogonality of the Fourier terms, it is established that the impact of each term assists in fitting the specified function closer to the scatter of the data, and therefore more closely approximate the underlying unknown function.

The Fourier offers a global approximation as it can exactly approximate any well-behaving unknown function when a linear combination of an infinite number of trigonometric terms is added, this way, the error in specifying the functional form – *specification error* – is reduced or even eliminated. However, adding an infinite number of Fourier terms is impractical as the coefficients of these terms will have to be estimated using an infinite sample (Mitchell and Onvural, 1996). That is why Gallant (1981) suggests truncating the Fourier series and accepting a given level of approximation error. Nonetheless, the potential approximation error of the Fourier form is further mitigated if the functional specification includes Fourier terms along with a second-order polynomial explanatory variable as represented by the Translog terms (Gallant, 1982). This results in the Translog being nested in the Fourier form. This combination will be referred to as the Fourier Flexible functional form.

The Fourier Flexible form Fourier is widely applied due to its capacity to approximate the underlying function over the whole range of the data (Gallant, 1982). The greater the number of parameters in a functional form to estimate, the more flexibility it has to approach the true unknown function as close as possible (Gallant 1982, Mitchell and Onvural 1996, and Fenn et al 2008).

The Fourier form is found to better fit banking data as opposed to the Translog given its considerable flexibility to approximate unknown functions as demonstrated by Mitchell and Onvural (1996), and Berger et al (1997), and Humphrey and Vale (2004). The latter studies all stress the need for using a more flexible cost functional form than Translog. For instance, using US and Norwegian banking data, McAllister and McManus (1993) and Humphrey and Vale (2004) observe misleading estimates when using Translog form, since it imposes a symmetric U-shaped average cost curve on the data, whereas the Fourier form permits for more curvature hence allowing for the possibility for an M-shaped curve as found by the two studies.

Girardone, Molyneux and Gardener (2004) confirm earlier studies such as Mitchell and Onvural (1996), Berger and De Young 1997, and Berger and Mester (1997) in that the standard U-shaped Translog specification is significantly restrictive and should be combined with the Fourier series. The authors explain that adding Fourier terms to the Translog combines the stability of Translog for data near the sample average, and takes advantage of the Fourier's flexibility which can accommodate for data that is scattered far from the mean. The result is a cost function that fits the entire rang of the data more flexibly. Altunbas et al (2000) argue that using the Fourier functional form – which is semi-parametric specification – avoids the inherent assumption of parametric specifications such as Translog in that the bank industry's true cost function has the Translog form.

The greater the number of Fourier terms the more flexible the function becomes, i.e. the closer it is to the true scatter of the data. The trigonometric transformation assists in accommodating for heterogeneity in bank sizes as Fenn et al (2008) explain, simply because this can increase the scatter of the data. McAllister and McManus (1993) and Mitchell and Onvural (1996) indicate that adding the Fourier terms to the Translog function is capable of correcting the bias of the Translog as it will be nested in the Fourier function.

As for truncating the Fourier series, White (1980) indicates that although an infinite number of Fourier terms will produce a perfect fit for an unknown function, a wellinformed truncation of the Fourier series can still provide a good approximation to the function over the whole data range.

This research follows White's argument and truncates the Fourier series for the profit and cost frontiers estimated in the first empirical chapter at the 1st order terms as tests revealed that truncating the series at the 2nd order terms will violate the concavity condition of the cost function. As for estimating, the Fourier series are truncated at the 3rd order terms following Berger and Mester (1997), Girardone, Molyneux and Gardner (2004), and Fenn et al (2008). A fourth-order Fourier terms has not been considered in the literature to date, however this is not necessarily suggestive of the incapability of higher order Fourier terms to improve the functional form fit. However, in relation to this research, the number of observations available was the main reason why a 4th order Fourier terms was not incorporated.

The appropriateness of applying this semi-parametric functional form is tested for using the log-likelihood tests as will be shown in both empirical chapters. Findings reinforce the view of Gallant (1982), McAllister and McManus (1993) and Mitchell and Onvural (1996) in that the Fourier provides a considerably better functional form fit. Testing the models' fit to the data by examining different specifications is believed to considerably reduce biasness in estimates because of reducing the possibility of specification error, and because the flexibility of the Fourier allows the data under study to reveal how close the specification is to the true unknown functional form. Mitchell and Onvural (1996) emphasize that increasing the number of parameters increases the variance of the test statistics when testing the null hypotheses[25], hence enhances the test power as the probability of making Type II error[26] becomes less. However, the number of Fourier terms included certainly needs to be sample-tailored as the problem of degrees of freedom should be in mind. This will extensively be discussed in the following empirical chapters.

3.2.3 Regulatory Conditions

[25] The null in a Wald or Log-likelihood Ratio tests states that the introduced terms are mutually insignificant.
[26] Type II error occurs when the null is accepted where in fact it is false. Reducing the probability of Type II error translates into lower probability for misperceiving the functional form, i.e. smaller specification error.

Verifying the regulatory conditions of the cost function ensures a well-behaving cost function which is significant to enhancing the reliability of the cost function estimation results. Therefore, consistent with production theory, the cost function should meet certain regulatory conditions so that its duality to production function (i.e. the transformation to production function) is maintained. Following Kim (1985) and Diewert and Wales (1987), these are: homogeneity, symmetry, positivity, monotonicity, and concavity.

1. Homogeneity implies that the cost function should be restricted to being homogenous of degree one in all input prices (that is achieving first order or linear homogeneity). Put simply, if all input prices increase by a factor λ, the level of total costs should increase by a factor of λ^1. This implies that a change in all input prices yields a proportionate change in Total Cost TC (unit elastic) which explains why the cost function is homogenous of degree 1 in all input prices as suggested by microeconomic theory.

2. Symmetry implies that second-order inputs and output own-interactive parameters should be symmetric as it is a necessary condition to achieve duality. Cost function symmetry is a result of applying the linear homogeneity property.

Linear Homogeneity and Symmetry are imposed on the Translog cost function, whereas other reminder three conditions are checked for.

3. **Positivity** implies that the cost function has to be positive in output quantities (that is marginal cost of outputs being positive $MCy_i > 0$), meaning that producing 1 extra unit bears a positive incremental cost or marginal cost regardless whether the marginal cost has an increasing or decreasing pattern.

4. **Monotonicity** implies that the cost function should be monotonic (increasing) in input prices, that is, costs are supposed to increase as input prices do. This is verified using Shephard's Lemma under which input demand functions are derived from the total cost function, yielding share input factor equations (e.g.

share labour and share capital) that are then checked for being non-negative.

5. **Concavity** – cost function should be concave in input prices, implying that if an input price increases, *ceteris paribus*, total costs will increase less proportionately at a given marginal rate of substitution – i.e., the expensive input is assumed to be substituted by the cheaper input at a given degree of substitution. Diewert and Wales (1987, p 47) suggest that a necessary and sufficient condition for the cost function's concavity is that the logarithmic second order derivatives of the Translog cost function w.r.t input prices should be related to its ordinary first and second order partial derivatives, such that:

$$\frac{\partial^2 \ln C(y,w,t)}{\partial \ln w_m \ln w_n} = \frac{\delta_{mn} w_m C_m}{C} - \frac{w_m w_n C_m C_n}{C^2} + \frac{w_m w_n C_{mn}}{C} \tag{20}$$

Where: $C_m \equiv \partial C(y,w,t)/\partial p_m$ stands for the ordinary first order partial derivative of the Translog cost function, and $C_{mn} \equiv C_{mn}\partial^2 C(y,w,t)/\partial p_m p_n$ is the second order partial derivative, and $\delta_{mn} = 1$ if $m = n$ and $\delta_{mn} = 0$ otherwise, and $C \equiv C(y,w,t)$. The left-hand side of the equality above is equal to the symmetric matrix of the estimated coefficient of the second-order input price term $\frac{1}{2}\sum_{m=1}^{n}\sum_{n=1}^{n}\gamma_{mn} lw_1 lw_2$, γ_{mn}, where the symmetric matrix of γ_{mn}, $A \equiv [\gamma_{mn}]$, should be negative semi-definite. Therefore, concavity can be achieved if $A \equiv [\gamma_{mn}]$ is negative semi-definite.

The regularity conditions are applied to the cost functions utilized in both empirical chapters and are thoroughly checked for. Satisfying these conditions is the criterion used to determine the choice of the preferred cost functions' specification that is estimated in both empirical chapters.

3.2.4 Structural Test: The Log-likelihood Ratio

This concept of LR tests is similar to that of *F* test which measures the increase (decrease) in the sum of squared residuals resulting from dropping (adding) variables from (to) a linear model yielding two different regression R-squared for the restricted and unrestricted models.

The test compares the values of the log-likelihood functions for the restricted and unrestricted models, computes the p-value of the resulting likelihood ratio test statistic, and therefore assesses whether the difference between the log-likelihood functions of the unrestricted and restricted models is statistically significant. The test uses the MLE estimation to maximize the likelihood function of a given model and produce its corresponding log-likelihood value. Adding new variables to the restricted model would normally lead to a larger value of the log-likelihood function, yet it remains to observe whether this increase in the log-likelihood value is statistically significant to decide whether the introduced variables have any statistically significant impact. The test statistic of the log-likelihood ratio is computed as twice the difference between the two log-likelihood values (Wooldridge, 2003):

$$LR = 2(L_{ur} - L_r) \tag{21}$$

Where L_{ur} and L_r are the log-likelihood values for the unrestricted (general) and restricted (nested) model respectively. Assuming $L_{ur} \geq L_r$ the LR ratio is therefore nonnegative. The test statistics LR has an approximate chi-square distribution under the null hypothesis such that $LR \sim \chi^2_\alpha$ with degrees of freedom *df* equal to the number of incorporated variables (restrictions or constraints) into the restricted model to be tested. The LR equation in (21) can be expressed differently as in Girardone, Molyneux, and Gardener (2004):

$$LR = -2\{\ln[L(H_0)] - \ln[L(H_1)]\} \tag{22}$$

The resulting LR in equation (21) and equation (22) is identical as both deliver nonnegative LR. $L(H_0)$ and $L(H_1)$ stand for the values of the log-likelihood functions under the null and alternative hypotheses. The H_0 postulates that the total impact of the added variables to the model is statistically insignificant, specifically, LR = 0 at $\alpha = 0.01$. The alternative hypothesis H_1 states that LR > 0, implying that the application of a one-sided chi-square test $\chi^2_{0.01}$ for the LR ratio.

Wald Chi2 test is also used along with LR test. Wald Chi2 can assess the goodness of fit of a given model by comparing its fit to the data including specific predictors to the fit of the same model excluding all predictors. Rejecting the null in this case suggests that regressors are statistically significant in explaining the independent variable's variations. It can also be employed to examine a constrained from of a given model against an unconstrained form of which. The null in this case states that there is no significant difference between the estimated parameters of the constrained and unconstrained models. Wald statistics and the corresponding p-value are calculated to assess the null hypothesis on the basis of: (1) the difference between the two models' parameter estimates and (2) the curvature of the log-likelihood function as measured by the second derivative of the function. This way Wald test evaluates the fit of the restricted against the unrestricted models (Greene, 2000).

It is worth noting that Wald chi2 and log-likelihood ratio tests can be used interchangeably as both examine the improvement or deterioration in the models fit by testing the constrained and unconstrained models. Both of these tests are applied as a robustness check for the different functional form specifications used in the two empirical chapters.

3.2.5 Frontier Estimation Techniques

Production theory suggests that the estimation of inefficiencies demands a full knowledge of the underlying function for an output-maximizing bank. This production function is unknown (Kim, 1985). Therefore, cost and profit functions are estimated instead since banks also have the economic objective of profit-maximization and cost-minimization. This advocates the estimation of the bank profit and cost functions, rather than the production function. Farrell (1957) therefore suggests the estimation of these functions utilizing the sample data by employing parametric or non-parametric techniques.

The efficiency literature suggests two main estimation methodologies: Parametric and Non-Parametric approaches. Briefly, parametric approaches specify a functional form to be estimated and attribute deviations from the efficient frontier to both inefficiency effect and random noise, while non-paramedic approaches avoid specifying a functional form and attribute deviations from the frontier entirely to inefficiency. The following paragraphs provide a brief introduction into efficient frontier estimation. This is followed by an extensive review of the estimation techniques.

Efficiency measurement assumes that producers seek the economic objective of production maximization, cost minimization, and profit maximization. Econometric techniques applied to gauge firms' efficiencies build on this notion of optimization to estimate production, cost, and profit function parameters using regression techniques. Deviations of the observed sub-optimal choices of some producers in the sample are caused by producing less output, incurring more costs, or making less profits than their efficient peers that are operating under similar conditions. Further, econometric estimation techniques should allow deviations from the efficient frontier to accommodate for the effects of: (1) the failure to optimize under similar conditions i.e. inefficiency, and (2) the influence of random effects (Fried, Lovell and Schmidt, 1993).

Frontier estimation is conducted according to different frameworks that can be distinguished on the basis of their underlying assumptions. Briefly speaking, the aim of these estimation methodologies is the identification of a given institution's areas of operational deficiencies by separating the latter from the random effects affecting its attainable profits or incurred costs according to parametric technique, while ascribing these deficiencies entirely to operational inefficiency according to non-parametric techniques which assume away any impact of possible random effects.

Calculating bank-specific efficiency scores requires comparing the observed performance of sample banks (i.e. banks off the frontier) with that of best performing sample banks, i.e. those situated on the efficient frontier that are operating under similar production technologies (i.e. inputs and outputs combinations given input prices and other factors).

It is argued that such an econometric approach to efficiency analysis is more effective than traditional benchmarking techniques such as financial ratio analysis, as the latter lacks the powerful optimizing methodology of the former. A number of studies have shown that the traditional financial ratio analysis to efficiency assessment is inferior to the econometric analysis of efficiency. Berger et al (1993) for instance argue that merely relying on financial ratios may produce misleading results as they either ignore the bank's product mix or input prices that are accounted for by frontier analysis, not to mention that they largely overlook incorporating elements of the economic environment that banks operate in.

At a macroeconomic level, the results of efficiency analysis help in identifying efficient banks in the banking system and highlight the factors that are driving the sub-optimal performance of other banks (Berger and Humphrey, 1997). Thus, investigating bank-level inefficiencies is important because it highlights areas of underperformance (i.e. spots waste in resources, excess costs, or deficiencies in revenues). Efficiency analysis also has important policy implications at a macroeconomic level as it assists policy makers is taking necessary steps to improve the operational efficiency of a given banking system.

Turning to the techniques used in estimating the efficient frontier, banking efficiency literature (such as Mester 1996, Berger and Humphrey 1997, and Berger and Mester 1997) postulates the following non-parametric approaches, including (1) DEA – Data Envelopment Analysis and (2) FDH – Free Disposal Hull, and three main parametric approaches, including: (1) SFA – Stochastic Frontier Analysis, (2) DFA – Distribution Free Approach, and (3) TFA – Thick Frontier Analysis. The following sections will elaborate on these approaches and will conclude by reflecting on the arguments supporting this research's choice of applying SFA.

3.2.6 Non-Parametric Techniques

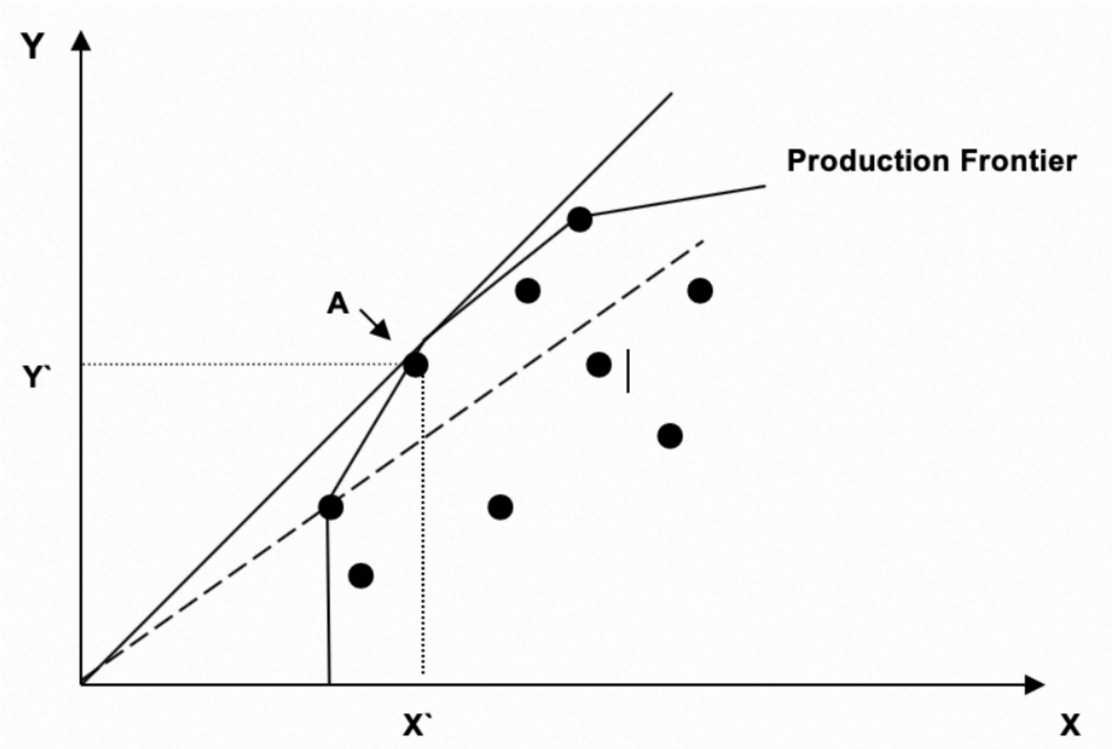
The literature acknowledges two main non-parametric approaches: Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH) (Berger and Mester, 1997). The generic property distinguishing these estimation methodologies from parametric approaches is that they do not specify a functional form for the efficient frontier. In essence, nonparametric approaches apply linear programming to local points to generate a benchmark (efficient frontier) that embodies the optimal cost incurred (for a cost-minimizing bank) or optimal profits achieved (for a profit-maximizing bank). The entire deviation of each bank from this optimal benchmark is attributed to operational inefficiency as the effects of random noise are assumed away.

3.2.6.1 Data Envelopment Analysis DEA

Pioneered by Charnes, Cooper, and Rhodes (1978), DEA distinguishes the best-practice observations or frontier banks according to their ability to produce highest possible level of outputs given inputs, or utilize the least inputs to produce a given level of outputs, under similar conditions. Technically speaking, DEA forms the efficient frontier by connecting the set of best-practicing banks in a piece-wise linear combination fashion that results in convex input/output possibilities which envelops data points.

All sample observations are enveloped by the piece-wise linear convex isoquant as shown in Figure 3 below such that observations could only lie on the frontier or to the right of which. Such technical property enables DEA to estimate the underlying function through the specification of the observed set of possibilities without being confined to a specific functional form (Drake and Howcroft, 1995 & Berger and Humphrey, 1997).

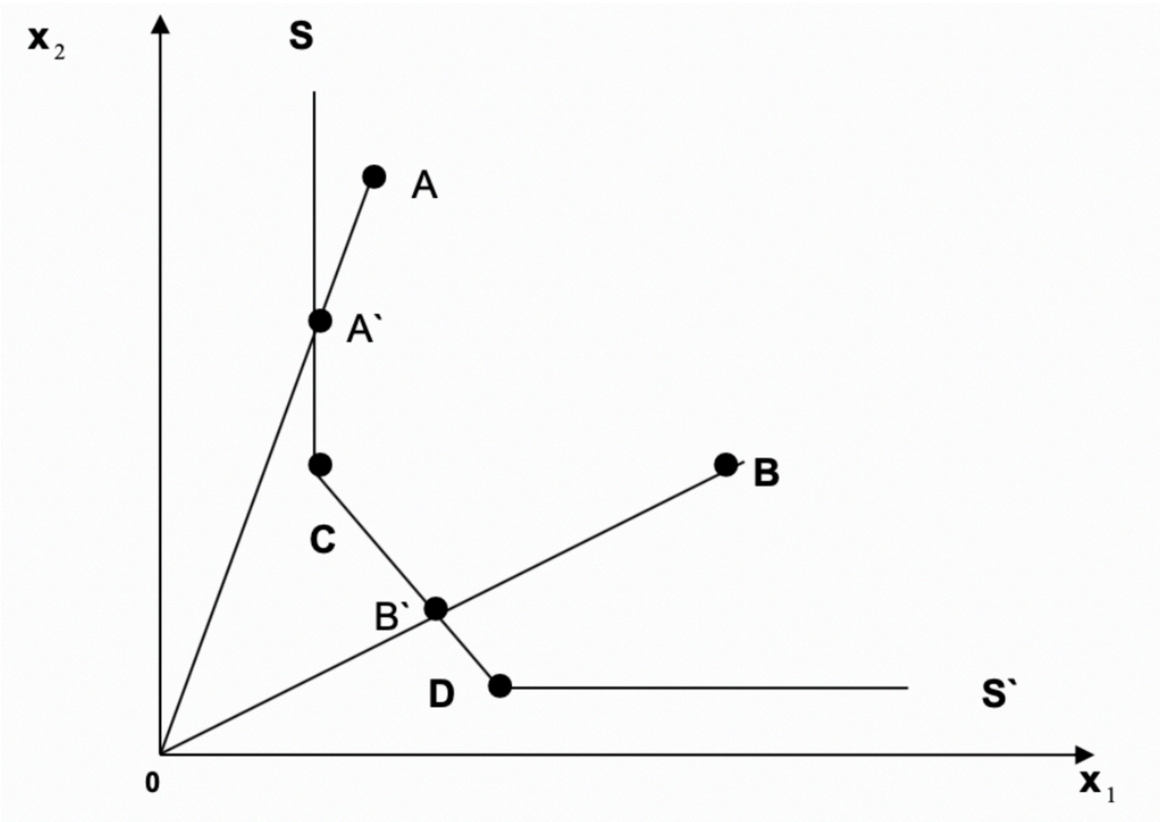
Figure 3: Piece-wise linear efficient frontier



The piece-wise efficient frontier envelopes all sample data such that no observation lies to the left of the frontier. The above depicted frontier represents a production frontier. It could similarly represent a profit frontier for instance by altering notations on the axes from output and input quantities to revenues and inputs prices respectively. The slope of the production frontier (y/x) at a given point yields the productivity achieved at that point, hence any efficient observation beyond the optimal scale (A) will have a marginally decreasing productivity as input increases.

Figure 4 shows how DEA is can provide individual efficiency scores. Bank-specific efficiencies are measured relative to the constructed frontier or surface (Coelli et al, 1998). The graph depicts a DEA piece-wise linear isocost (ss') that is associated with output level y and input prices x_1 and x_2. Producers or sample observations (or DMUs - decision making units in DEA terminology) located at points C and D define the efficient frontier. As was illustrated in section 6.1.2, technical inefficiency at point A is defined by the ratio of $0A'/0A$. The same output level achieved by producing at point A` can be realized by producing at point C with less of input $x_2$ being used, therefore it is questionable whether producer A` is technically efficient although it is situated on the efficient DEA frontier. This issue highlights the problem of input excess or input slacks (Battese and Coelli, 1998) which indicates that any non-zero input slacks should be reported along with DEA technical efficiency scores.

Figure 4: Efficiency Measurement under DEA (Coelli and Battese, 1998)



DEA imposes no functional form on the data but it ignores data noise, and that is why it requires no distributional assumptions for inefficiency. Relaxing these constrains permits the shape of DEA frontier to be flexible, yet it prevents further statistical analysis to be applied on the DEA frontier (Bauer, 1990). Therefore, since DEA is a nonparametric technique, statistical hypothesis tests are difficult to conduct. In this context, conducting Log-Likelihood ratio tests for instance to observe the potential improvement that the introduction of risk measures would have on the efficient frontier would be impossible since DEA defines no functional form. Consequently, as the main objective of this research is to modify the efficient frontier's functional form, this has led to the exclusion of DEA technique.

Moreover, the dismissal of the random error component hampers the validity of non-parametric models especially when it comes to dealing with cross-country banking data. This is because sources of the random error component are, according to Berger and Humphrey (1997); (1) measurement errors, (2) luck that may boost a bank's performance in a given year relative to the next, and (3) the inaccuracies and discrepancies of accounting rules. Therefore, should any of these factors exist, inefficiency scores will most likely to be biased (inflated or deflated) as they would inherently include the random error effect if not isolated according to Berger and Humphrey (1997) and Battese and Coelli (1998), this is because the entire deviation from the frontier will be ascribed to inefficiency, whereas in fact it is not all attributed to inefficiency. For this reason, amongst others, non-parametric approached were not used in this research. Moreover, since this research uses accounting data, it is therefore important to know if the random error component exists in the first place. In fact, the data is tested for stochasticity, and strong evidence of significant stochasticity is found in the two-sample data used in this research. Stochasticity simply suggests that the random error term in the data exists and is statistically significant.

Attempts to improve DEA model has in one aspect resulted in the development of bootstrapped-DEA. Some research, Banker (1993) Grosskopf (1996) and Berger and Humphrey (1997), suggest that the deterministic nature of DEA can be relaxed by introducing a re-sampling technique such as bootstrapping. Bootstrapping can approximate the distributions of estimated efficiencies, and therefore standard errors can be obtained accordingly. Such transformation would enable the interpretation of DEA estimates as a maximum likelihood estimate. However, very few studies have followed this route to DEA estimation (Cummins et al, 2003 for example) due the need for careful specification of the data-generating process which can produce arbitrary distributions as Simar and Wilson (1995) explain.

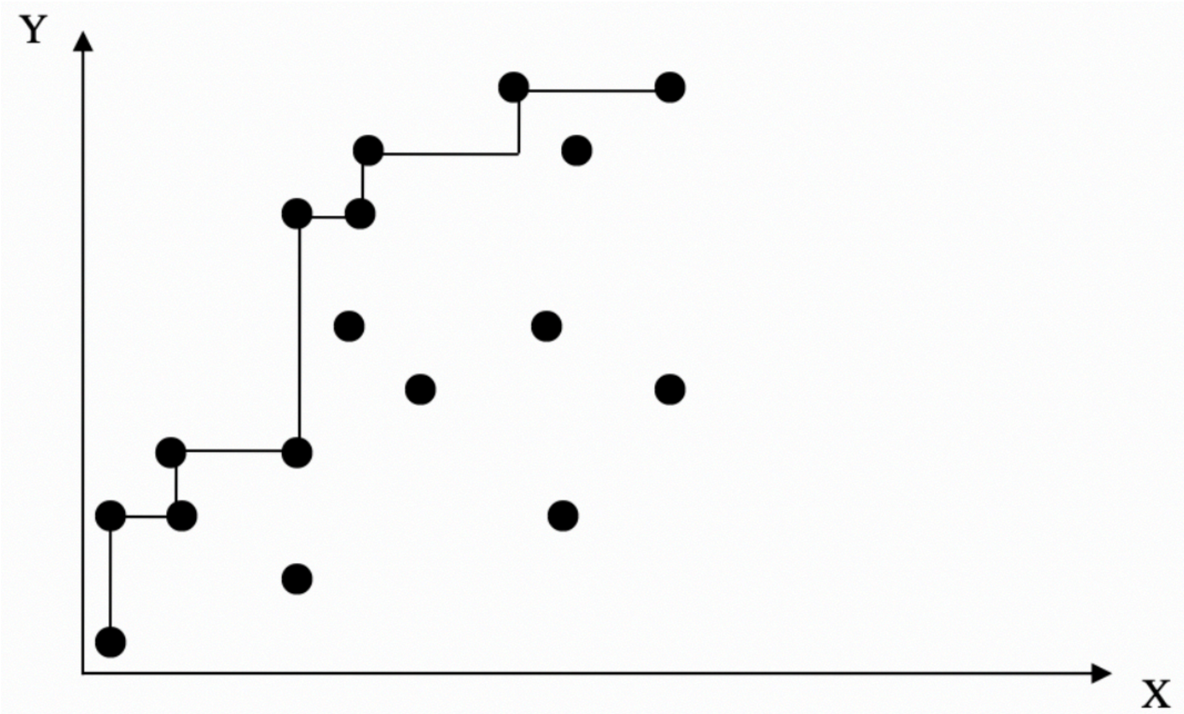
Further advances also involved shifting the deterministic nature of DEA to accommodate for the data noise. This has resulted in the emergence of the so-called Stochastic DEA or SDEA developed by Li (1998). The generic aim is to adjust DEA for random errors. This is achieved by allowing a certain probability to violating normal DEA constraints. The constraints, representing the relationship between an individual decision-making unit (DMU) and the frontier DMU, are made more difficult to satisfy. This has the effect of raising the measured technical efficiency of every DMU relative to the deterministic model. This shifts the SDEA frontier closer to where the data is most concentrated. This entails that, the smaller the probability that a single constraint will be satisfied for an observation, the closer to the frontier local point would be, implying that the average SDEA efficiency scores will never be lower than those of DEA (Sarafidis, 2002, p 20).

The problem with the SDEA model lies in the subjectivity of the process of associating random probabilities to the different DEA constraints, which in turn makes the model's results very sensitive to this assumption (Sarafidis, 2002). Brazdik (2005) explains that, for this reason, the application of SDEA has been relatively limited. Lastly, the application of the bootstrapped-DEA or SDEA approaches to this study is deemed to add little value, given that the core purpose of this research is to advance efficiency analysis through the incorporation of risk, rather than testing the impact of employing different estimation methodologies on efficiency estimations.

3.2.6.2 Free Disposal Hull FDH

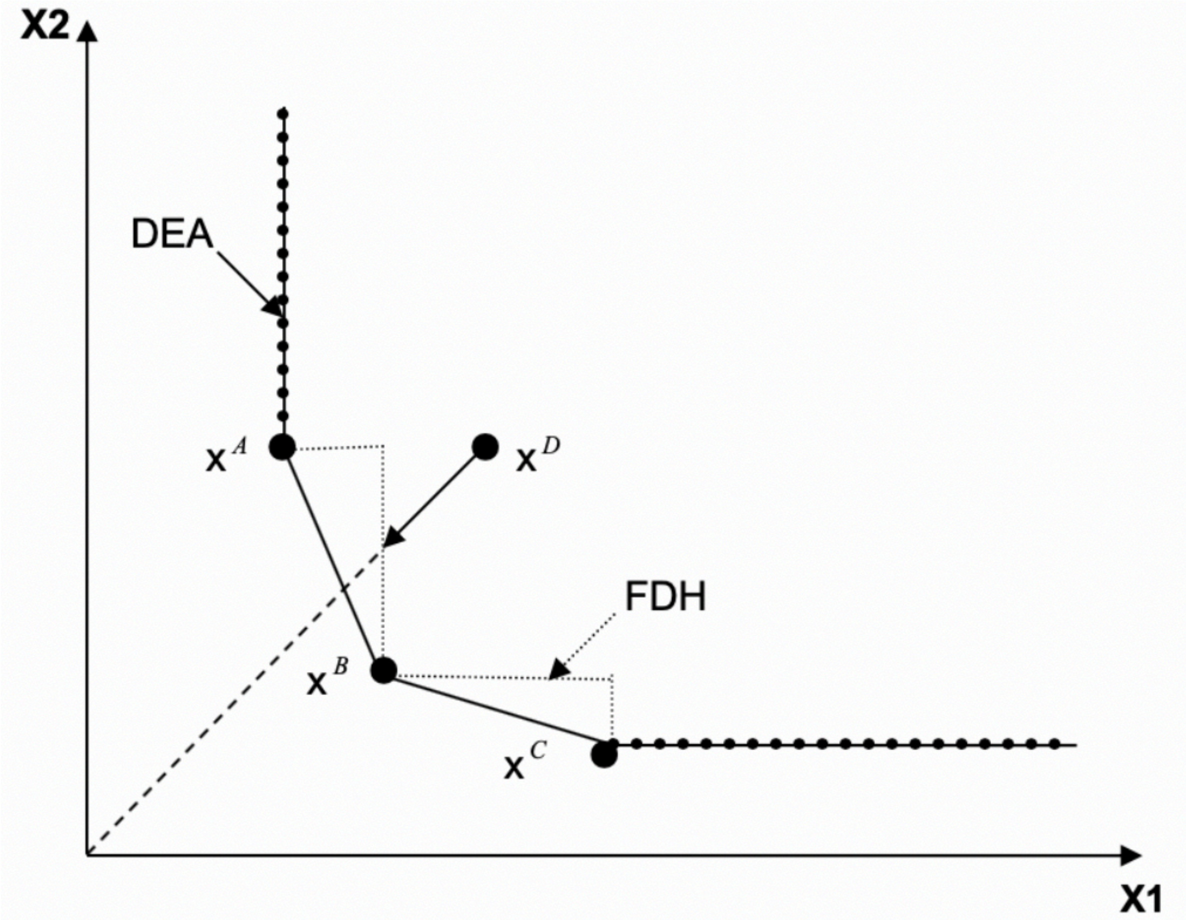
FDH was developed by De Prins, Simar & Tulkens (1984). FDH is a special case of DEA model as it follows a linear programming methodology to estimating the frontier. FDH forms the efficient frontier which envelops the observed data in a step-wise process as shown in Figure 5.

Figure 5: Step-wise efficient production frontier



The distinctive shape of the FDH frontier is partly a result of not imposing the convexity assumption on the production function as does DEA. Therefore, observations labelled as inefficient with FDH will also be under DEA but the reverse is not certain as demonstrated by Figure 6. That is because points connecting DEA vertices are not part of the FDH frontier as shown below. In this context, the FDH generates the production possibilities set in a manner that envelopes DEA vertices with the FDH points being interior to these vertices (Berger and Humphrey, 1997).

Figure 6: Efficiency Measurement under FDH (Fried et al, 1993)



DEA allows for the possibility of linear substitution between observed input combinations as represented by the different points on the isoquant (which is generated from the observations in piecewise linear forms), whereas FDH discards this possibility by allowing the isoquant to look as a step function that is constructed by intersecting different linear combinations of the local (observed) inputs (Berger and Humphrey, 1997, p 5). However, this creates a more serious problem of input slacks (excess) than in DEA. Moreover, as can be observed in Figure3, FDH frontier surrounds the data more tightly than DEA, which results in technical efficiency estimates of FDH to be – on average – smaller than those of DEA (Tulkens, 1993), since deviations from the frontier will be smaller under the FDH.

All in all, FDH and DEA suffer a major drawback since they assume away the random error when estimating the efficient frontier, accordingly, non-parametric techniques imply – as Berger and Humphrey (1997) explain – that there is: no measurement error in estimating the frontier, no luck that may boost the estimated performance of the observed unit from one year from the next, and no inaccuracies or discrepancies of accounting rules in measuring outputs and inputs.

As the random error is tested for stochasticity, and the results clearly demonstrate that the panel's error term is significantly stochastic hence suppressing the random error would yield inefficient (bigger standard errors) and biased estimates (since the mean of the estimated coefficient's distribution is very likely to deviate from the true parameter β due to the ignored impact of the estimation error term).

To conclude, this research dismisses non-parametric approaches for two reasons: (1) because they ignore the possible impact of statistical noise or random error, and (2) because they do not allow for specifying a frontier functional form and therefore there is no way to assess the potential impact of the risk factors proposed in this research. Drake et al (2006, p 1444) sum up the argument against non-parametric methods by highlighting their major drawback in that "there is no standard statistical test to determine whether the researcher has utilized the correct set of non-controllable inputs or outputs". Hence, without specifying a functional form, this research will be deviating from one of its main objectives as it is quite difficult to decide whether a certain set of risk factors have improved the fit of the model to the data or not.

3.2.7 Parametric Techniques

Contrary to non-parametric methodologies, parametric approaches specify a functional form to be estimated and make clear assumptions about the distributions of the inefficiency and random error components. The literature acknowledges three main parametric techniques including: Distribution Free Approach DFA, Thicker Frontier Approach TFA, and Stochastic Frontier Approach SFA. The discussion begins by exploring the DFA and TFA methods and concludes with methodology applied in this research, SFA, along with the rationale for adopting SFA.

3.2.7.1 Distribution Free Approach DFA

Introduced by Berger (1993), the DFA requires a panel data set and is capable of providing bank-specific efficiency estimates. It assumes an estimation composite error term that comprises of two components: a random term and inefficiency term. However, DFA makes two strong assumptions upon estimating the efficient frontier.

First, is that the estimated errors or residual ($\hat{\varepsilon}_{it} = \hat{v}_{it} + \hat{u}_{it}$) are assumed to average out over time, hence bank-specific residual becomes constant such that $\hat{\varepsilon}_i = (1/T)\sum_t \hat{\varepsilon}_{it}$. Second, is that the random component of the residual is assumed to cancel out over time, therefore the entirety of the residuals will constitute for inefficiency terms only, i.e. $\hat{\varepsilon}_i = (1/T)\sum_t \hat{\varepsilon}_{it} \cong \hat{u}_i$. This implicitly implies that the DFA defines inefficiency as time-invariant.

The DFA does not make any distributional assumptions about the inefficiency term, but assumes that the inefficiency term (u_i) is a random variable distributed independently from the model's regressors. The DFA has an inherent assumption in that bank-level efficiency is consistent over time (systematic time progression trend), whereas the random error terms are more likely to cancel out one another over time (average to zero) given a panel data set specifically (Bauer et al, 1998).

Technically, bank-specific cost inefficiency, for instance, is estimated as the difference between the average residual of a given local point (observed bank) and the average residual of the efficient bank – that is banks with the smallest residual value $\min_i(\hat{\varepsilon}_i)$ – assuming that the model's random error components will cancel out over time. Accordingly, cost efficiency is expressed as (Kumbhakar and Lovell, 2000, p 180):

$$\hat{CE}_i = \exp\{-[\hat{\varepsilon}_i - \min_i(\hat{\varepsilon}_i)]\}$$

(23)

Estimated efficiency is bounded between 0 and 1. A clear disadvantage of the DFA is the assumption that efficiency is time-invariant which undermines the credibility of DFA's estimates if the panel involves a long period of time. On the other hand, if the length of the panel is small, the assumption that the random component will average out to zero is more likely to be violated hence distorting the inefficiency estimates as the model's residual will not entirely be attributed to inefficiency (Kumbhakar and Lovell, 2000). This particular weakness of DFA prompted research into exploring the optimal panel length (T) for which the two assumptions would hold. De Young (1997) investigated this issue and found that the optimal value of T is the first value of T when the variance of which, $\sigma^2(T)$, stops decreasing. In any case, this suggests sacrificing some of the panel's observations which may result in incorrect conclusions and can create the degrees of freedom problem especially for relatively small samples.

Moreover, given the assumption that the random component cancels out over time, DFA accommodates for observations with extreme average residual values by truncating the distribution of the random error component at 1%, 5% and 10%. This implies that applying the DFA would require accepting a given level of degrees of freedom, which is problematic in case of relatively small samples as Berger (1993) illustrates. Therefore, the DFA approach mitigates the impact of data outliers because it truncates extreme average residuals (Berger and Humphrey, 1997). There are three different DFA approaches: DFA-Within, DFA-GLS, and DFA-Truncated. All of these approaches have a similar assumption in that inefficiency is fixed over time, however they differ in the way the estimation is conducted and by the way they treat the model's residual (Bauer et al, 1998).

To conclude, the application of DFA is discarded for different reasons. First, because it ignores the impact of the random error component in the sense that it is assumed to cancel out over time. Second, it ensures the 'cancelling out' of the random component by truncating the data at both ends, which in turn may create the degrees of freedom problem and can produce biased results. Third, the stochasticity test applied in this research to both data sets shows that the specified cost and profit functions demonstrate significant evidence of the model's error being stochastic, in other words, this effectively invalidates the assumption that the random error component of the composite error term cancels out as far as this research's sample data is concerned.

3.2.7.2 Thick Frontier Analysis TFA

Developed by Berger and Humphrey (1991, 1992), the Thick Frontier Analysis TFA is a less structured frontier estimation approach compared to the DFA but can be applied to cross-section and panel data settings as Kumbhakar and Lovell (2000) explain. The following discussion illustrates the TFA from a cost efficiency perspective and in a cross-sectional framework. Technically speaking, TFA requires calculating point average costs for the sample data and then stratifying the data in a descending manner in terms of the average costs calculated. This results in forming average cost quartiles. Accordingly, the group of banks situated in the bottom quartile are assumed to be the most cost efficient and constitute the thick frontier, and banks located in the top quartile are considered as the least cost-efficient (as they have the highest average costs) relative to the thick frontier.

Having classified the data as such, a separate cost function for the top and bottom quartiles is estimated. TFA relies on the following assumptions: (1) within each quartile, residual variations are assumed to represent statistical noise only, and (2) variations in predicted average costs between top and bottom quartiles are assumed to account for cost inefficiency in the top quartile. In other words, because cost functions are estimated separately for top and bottom quartiles ($Q1, Q4$), differences between the estimated parameters (β^1, β^4) are assumed to reflect differences in cost efficiency between banks in the top and bottom quartiles ($Q1, Q4$). Kumbhakar and Lovell (2000) argue that these two assumptions may not hold as such, meaning that the TFA is likely to produce inaccurate estimates, but it is a useful tool that gives an idea for the possible magnitude of inefficiencies.

The average cost inefficiency CI_i in cost inefficient quartile $Q4$ can therefore be expressed by the following relationship (Kumbhakar and Lovell, 2000, p 178):

$$CI_i = \frac{[c(y_i, w_i; \beta^4)/y_i] - [c(y_i, w_i; \beta^1)/y_i]}{[c(y_i, w_i; \beta^1)/y_i]}$$

(24)

Where CI_i represents increase in average costs brought about by the inefficient producer as a percentage of the average cost of the efficient producer located in quartile $Q1$.

What is more, the TFA does not postulate any distributional assumptions regarding the estimated model's error term or residual. Alternatively, it attributes inefficiency to differences between estimated parameters in the top and bottom quartiles as equation (24) shows. This in fact could potentially produce biased and inconsistent efficiency scores because of ignoring statistical noise, because the impact of exogenous factors such as measurement errors and luck and possibly omitted variables are not separated from inefficiencies. It follows that the TFA provides no distributional assumptions related to the error and inefficiency terms as a way to separate them since it ignores the impact of statistical noise in the first place. However, the main advantage of TFA is that it mitigates the outliers-effect as a result of sample stratification procedure.

More seriously, the TFA does not provide bank-specific efficiency estimates, rather, it produces one cost efficiency estimate for a 'hypothetical' mean firm in the high-cost quartile relative the hypothetical mean firm in low-cost quartile (Berger and Humphrey, 1997 and Bauer et al., 1998). Therefore, the TFA measures the 'core' efficiency for the entire sample rather than providing an observation-specific efficiency score. Lastly, TFA does not include all the sample data into the analysis, it rather considers 50% or 40% of data included in the top and bottom quartiles, since inefficiencies are determined as the difference between average costs in the top and bottom quintiles. This may lead to a serious problem of degrees of freedom and can produce biased results since not all observations in the sample data are considered. For these reasons, the TFA was discarded in this research and the Stochastic Frontier Approach SFA is applied instead. The SFA is the subject of the next section.

3.2.7.3 Stochastic Frontier Analysis SFA

First proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and Van Den Broeck (1977), SFA is an econometric modelling method that requires specifying a 'stochastic' functional form such that it incorporates an additional random error component v_i into the model to accommodate for the impact of random factors or statistical noise on the dependent variable which can be the result of measurement errors, luck, and the effects of unspecified (omitted) variables.

The following discussion gives more insight into the evolution of SFA model given its central importance to this research. To start with, Aigner and Chu (1968) suggested the estimation of a Cobb-Douglas production function defined as:

$$\ln(y_i) = x_i\beta - \varepsilon_i \tag{25}$$

Where:

$\ln(y_i)$ is the logged-form of output for the i^{th} bank ($i = 1,2,...,n$),

(x_i) is input quantities utilized to produce the output,

(β) represents a set of parameters to be estimated,

(ε_i) is the non-negative random variable associated with the bank's technical efficiency.

Aigner and Chu estimate the inefficiency term under this specification as the ratio of the observed level of output for the i^{th} bank (y_i) to the estimated value of the frontier output [$\exp(x_i\beta)$] which is associated with no technical inefficiency.

Technical inefficiency can be expressed, in exponential terms[27], as follows:

$$TE_i = \frac{y_i}{\exp(x_i\beta)} = \frac{\exp(x_i\beta - \varepsilon_i)}{\exp(x_i\beta)} \Rightarrow TE_i = \exp(x_i\beta - \varepsilon_i) \times \exp(-x_i\beta)$$

$$\Rightarrow TE_i = \exp(-\varepsilon_i) \qquad \textbf{(26)}$$

Afriat (1972) and Richmond (1974) estimate the non-stochastic production function defined in (25) using the Maximum Likelihood MLE assuming that the inefficiency term follows the gamma distribution. Coelli et al (1998) state the disadvantage of estimating a deterministic (non-stochastic) model specified in (25) in that it ascribes all deviations from the frontier to technical inefficiency and ignores the potential impact of exogenous factors such as measurement errors, luck, and potentially omitted variables, hence it is likely to produce biased inefficiency estimates as a result.

Aiming to overcome this potential weakness of the deterministic model, Aigner et al (1977) introduced the stochastic frontier model, which is estimated using the stochastic frontier approach SFA, to more accurately accounts for noise in the data by disentangling the random error into two components: an inefficiency term and a random error term. The production frontier specified in (25) can therefore be expressed as a stochastic production function such that:

$$\ln(y_i) = x_i\beta + (v_i - u_i) \qquad \textbf{(27)}$$

Where:

- ($x_i\beta$) is the deterministic part of the stochastic functional form
- ($v_i - u_i$) is the composite error term
- (v_i) represents the effect on outputs due to exogenous influences. v_i is a set of identically independent and distributed i.i.d. random variables that follow and

[27] If $\ln(y_i) = \alpha$ then: $y_i = e^{\alpha}$.

are independent of the inefficiency terms u_i, hence $v_i \overset{i.i.d.}{\sim} N(0,\sigma_v^2)$. Berger and Mester (1997) refer to the random error as idiosyncratic error or statistical noise.

- $(-u_i)$ represents technical inefficiency. (u_i) is a set of inefficiency terms that are also assumed to be i.i.d. following exponential distribution, or half-normal distribution: $u_i \sim iidN^+(0,\sigma_u^2)$.
- $(x_i\beta + v)$ is the stochastic frontier

The stochastic property of the production model defined in (27) stems from splitting the estimated residual into random and inefficiency components (Coelli et al, 1998). Therefore, SFA model assumes that the residual (which constitutes the portion of the dependent variable's variation that cannot be explained by the model) represents the inefficiency effects after isolating the random effects caused by measurement errors and luck. These two components are separated using different distributional assumptions, where inefficiency effects are assumed to follow one-sided (usually half-normal) distribution, whereas the random effects are assumed to follow two-sided normal distribution. Bank-specific inefficiency is then calculated as the ratio of its predicted output of the observed bank to the predicted output of the best-practicing bank in the sample. This is graphically demonstrated by Figure 7 below.

Figure 7: The Stochastic Frontier Production Function (Coelli, 1998, p 180)

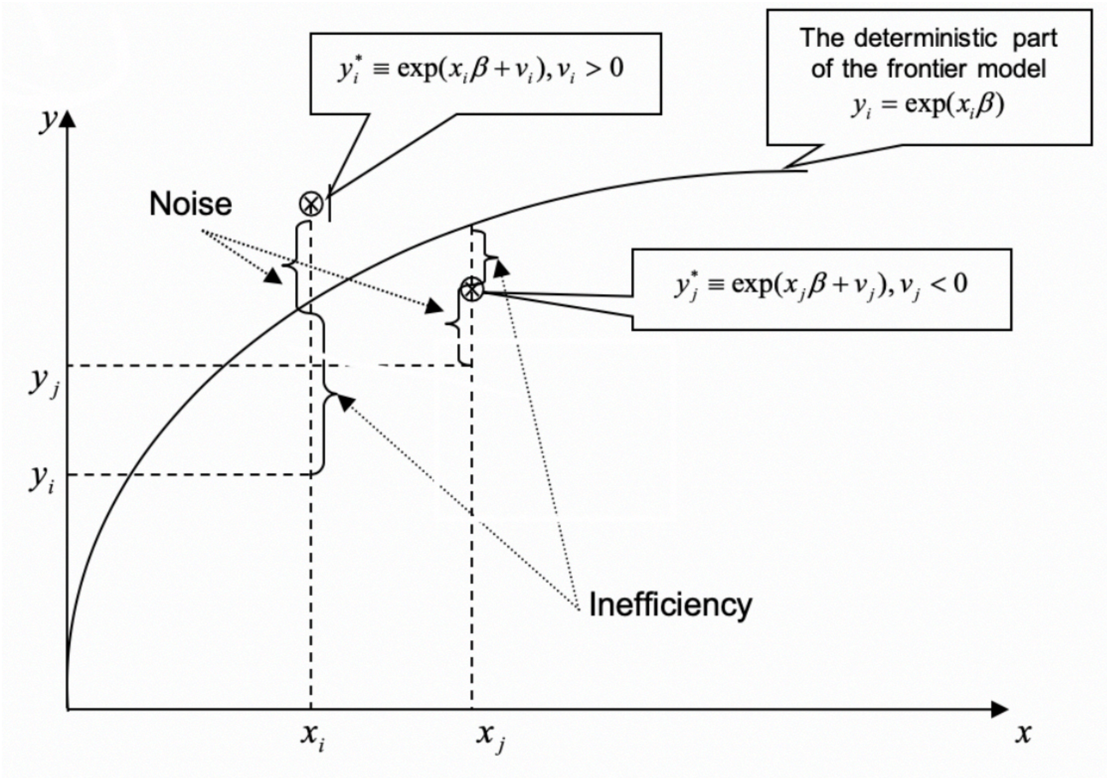


Figure 7 explains the main components of the stochastic production frontier where the horizontal and vertical axes represent input and output levels for two firms *i* and *j*. The shape of the frontier's deterministic component implies the microeconomic law of diminishing returns to scale. The observed input-output value for firm *i* and *j* are defined by the point (x_i, y_i) and (x_j, y_j) respectively. As can be seen from the graph, deviation from the frontier is attributed to statistical noise and inefficiency. Observed outputs may lie below or above its deterministic part conditional on the random error being negative or positive since it is assumed to follow a two-sided normal distribution $v_i \sim N(0, \sigma_v^2)$ according to Coelli et al (1998).

To produce bank-specific estimates, SFA first requires specifying a functional form to be estimated, such as Cobb-Douglas, Translog, and Fourier Flexible form…etc. The stochastic function is then estimated using the maximum likelihood method (MLE) to produce estimates for the model's parameters. Second, using these MLE estimates, bank-specific technical efficiency scores are produced by disentangling the estimated residual term into a statistical noise, and a technical inefficiency component (Kumbhakar and Lovell, 2000). So, for instance, estimating a cost function under SFA involves labelling "a bank as inefficient if its costs are higher than those predicted for an efficient bank producing the same input/output configuration and the difference cannot be explained by statistical noise" (Altunbas et al, 2000, p 1607).

The specification in equation (27) expresses a stochastic production frontier. Specifying a stochastic cost frontier simply requires changing the dependent variable to represent total costs instead of outputs, and the sign of the inefficacy component of the error term $(-u_i)$ to become $(+u_i)$ as suggested by Coelli (1996). Thus, the stochastic cost frontier can be expressed as follows:

$$\ln(TC_i) = x_i\beta + (v_i + u_i) \tag{28}$$

Where:

- (TC_i) is the total cost of production for bank i,
- (u_i) is a nonnegative random variable such that $u_i \overset{iid}{\sim} N^+(0, \sigma_\mu^2)$ and is assumed to account for cost inefficiency.

The cost inefficiency term (u_i) defines how far a given observation is operating above the cost frontier, hence $u_i \geq 1$. On the other hand, to express a stochastic (alternative) profit frontier, the definition of the random error term for will be similar to that of the production frontier such that $(v_i - u_i)$, with the independent variable being profits instead of outputs.

This can therefore be expressed as:

$$\ln(P_i) = x_i\beta + (u_i - v_i)$$

(29)

Profit inefficiency term (u_i) defines how far a given observation is operating under the profit frontier, hence $u_i \leq 1$.

To produce bank-specific inefficiency estimates, SFA takes each observation's error term ε in each time period in the panel and estimates the inefficiency score as conditional on the observed error. Moreover, efficiency estimates under SFA are consistent estimates as shown by Ferrier and Lovell (1990) and later by Fiorentino et al (2006) because SFA estimates are insensitive to the sample heterogeneity (dispersion around the mean values) or the effect of outliers. This is further illustrated in the following section.

3.2.7.4 The Rationale for Choosing SFA

On the rationale for choosing SFA as opposed to DEA, Vennet (2002, p 263) argues that nonparametric techniques, compared to SFA, generally produce less reliable estimates because they discard statistical noise in the data. Furthermore, The DEA has been empirically shown to produce inconsistent efficiency estimates due to its sensitivity to sample heterogeneity and to the effects of outliers. Fiorentino et al (2006) use a panel data se of 34,192 banking observations over 1993-2004 to test for the sensitivity of DEA and SFA efficiency estimates to sample heterogeneity and to the effect of outliers. To test the sensitivity to sample heterogeneity, the sample was structured (grouped) in three different ways: by year, by asset size, and by both criteria. The authors found that mean cost efficiency under DEA changed significantly under the different structures, suggesting that DEA estimates are significantly sensitive to sample heterogeneity, this is in contrast to SFA estimates which experienced no significant change given the different settings. This suggests that DEA estimates are inconsistent.

To test the effect of outliers on estimates, Fiorentino et al (2006) excluded for this purpose 24 observations (outliers) only out of a total of 34,192 and re-estimated the model under DEA and SFA. Results indicated that mean efficiency scores under DEA increased dramatically from 13% to 37%, whereas under SFA results were stable and experienced almost no change (Fiorentino et al, 2006, p 4).

Thus, the justification for applying SFA approach as opposed to DEA is threefold: (1) the need for specifying a functional form since this best serves this research's purposes, (2) DEA ignores accounting for a random error in the data whereas the data used in this research shows strong evidence of existing statistical noise, and (3) SFA seems to deliver more consistent and therefore more reliable since SFA is less sensitive to sample heterogeneity and the effects of outliers.

In relation to other parametric techniques, SFA is more appropriate than either TFA or DFA. It dominates the TFA because the latter is unable to produce individual bank efficiency scores which would hinder the fulfilment of this research objective. With regards to the DFA, although some research has shown a relatively high correlation between SFA and DFA rankings (Bauer et al, 1998), SFA is preferred to the DFA since the DFA implies a strong assumption under which the random component of the data noise is believed to average to zero. Ruling out the impact of the random component as such may inflate/deflate the average profit /cost inefficiencies over the panel period if this random component is found significant since average deviations will be totally attributed to inefficiencies. Berger (1993) affirms this view by arguing that some noise in the data might be persistent over time and therefore should be accounted for.

Moreover, contrary to SFA, Bauer et al (1998) refers to a potential loophole in the DFA methodology based on the somewhat an arbitrary assumption that inefficiency is time-invariant as it assumes that inefficiency tend to average out over time along with the random error component. This in fact rules out the influence of any factors that may change the level of inefficiencies over time such as technological advances or other environmental factors. The other weakness of the DFA stems from providing no assumptions on the probability distribution for the inefficiency and the random error component, which may be the result of the first assumption in that the random error portion of the noise cancels out over time. Therefore, SFA dominates other parametric techniques as it permits the decomposition of the model's residual into technical inefficiency and random components based on their different distributional specifications (Ferrier and Lovell, 1990). This property of SFA perfectly serves the purpose of this research.

3.2.8 Maximum-Likelihood Estimation MLE

The estimation procedure applied in this research (SFA) uses the concept of maximum likelihood estimate MLE to produce estimates for the specified in cost and profit frontier models. Estimating a set of parameters using the MLE aims at finding their estimates that maximize the corresponding likelihood function (Harris and Stoker, 1998). Parameters' estimates that maximize the likelihood function correspond to the lowest standard errors possible to these estimates[28]. A likelihood function is a probability or probability density function $f(x)$ representing the relationship between a random variable's different outcomes and their corresponding probabilities (Wooldridge, 2003). In the case of frontier function, the likelihood function represents the probability density function of the frontier's composite error term (the unknown variable) given the observed values of the models' parameters.

MLE is widely applied as it provides asymptotically efficient[29] estimator – i.e. MLE estimates are efficient for big sample sizes – if the model is correctly specified. As will be discussed in the following two empirical chapters, the preferred specification for the models applied are carefully tested for prior to performing estimations.

If the model to be estimated is found to be stochastic, Coelli (1996) and Wooldridge

---

[28] The lower the standard errors of estimates the closer that these estimates are to their true unknown values of the unknown population which the sample was drawn from.

[29] An asymptotically efficient estimator W of parameter θ suggests that it is efficient, unbiased, and consistent estimator.

- Efficiency implies how far the estimator is away from the true value of the estimated parameter θ, suggesting that, for two estimators, the more efficient one is that with the minimal mean squared error MSE. Assuming that W is an estimator of θ, then W is an efficient estimator of θ if $MSE(W) = E[(W-\theta)^2]$ (Wooldridge, 2003, p 740).

- Unbiasedness means that if the estimator's "probability distribution has an expected value equal to the parameter it is supposed to be estimating" (Wooldridge, 2003, p 735), that is: $E(W) = \theta$, hence $Bias(W) = E(W) - \theta$. The probability distribution of the estimator W represents the likelihood of its various outcomes resulting from drawing random samples. Obviously, unbiasedness does not imply that the estimate generated by a specific sample's data is equal to θ, instead, it implies that if an infinite number of random samples on x_i were to be drawn from the population and the estimate for each sample is calculated, then average of these estimates should converge to θ

- **Consistency** suggests that the estimator's probability distribution becomes more and more concentrated around the true parameter θ as the sample size increases, indicating that the estimator W will deviate less from θ for larger sample sizes (Wooldridge, 2003, p 741).

(2003, p 747) suggest using the MLE estimator as it tends to produce efficient (smallest variances), unbiased and consistent estimators (in large samples). To give more details on the MLE methodology, let us consider the following cost function:

$$\ln C_i = \ln c(y_i, w_i, t, \beta) + u_i + v_i$$

(30)

The estimation of the unknown parameters of equation (30) requires the following:

- Specifying the functional form (Cobb-Douglas, Translog, or Fourier Flexible) for the deterministic part of the cost function $c(y_i, w_i, t, \beta)$.
- Specifying the distributional assumptions of composite error components: the random variable (statistical noise) $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$ and the inefficiency term $u_i \overset{iid}{\sim} N^+(0, \sigma_u^2)$. The two components are also assumed to be distributed independently from each other, and from the model's regressors.

Accordingly, the MLE maximizes the likelihood function corresponding to the cost function specified in (30) using the underlying data. This implies that the maximum likelihood function needs to be specified. The following discussion shows the steps with which the log-likelihood function is derived according to Aigner, Lovell, and Schmidt (1977) & Kumbhakar and Lovell (2000).

The density function of two-sided normally distributed random variable component v_i is defined as:

$$f(v_i) = \frac{1}{\sqrt{2\pi\sigma_v}} \exp(-\frac{v_i^2}{2\sigma_v^2})$$

(31)

Also, the density function of the non-negative normally distributed u_i is defined as

$$f(u_i) = \frac{2}{\sqrt{2\pi\sigma_u}} \exp(-\frac{u_i^2}{2\sigma_u^2})$$

(32)

As the two components are assumed to be independent, the joint density function of both is the product of their individual density function, thus the product of equation (31) and (32) yields:

$$f(u_i, v_i) = \frac{2}{2\pi\sigma_u\sigma_v}\exp(-\frac{u_i^2}{2\sigma_u^2} - \frac{v_i^2}{2\sigma_v^2})$$

(33)

This joint density function can be then expressed in terms of the composite error term ε_i and its inefficiency component u_i, since $\varepsilon_i = v_i + u_i$ hence $v_i = \varepsilon_i - u_i$. This yield:

$$f(u_i, \varepsilon_i) = \frac{2}{2\pi\sigma_u\sigma_v}\exp(-\frac{u_i^2}{2\sigma_u^2} - \frac{(\varepsilon_i - u_i)_i^2}{2\sigma_v^2})$$

(34)

To obtain the marginal density function of the composite error term ε_i, the joint density function $f(u_i, \varepsilon_i)$ is integrated with respect to u_i yielding:

$$f(\varepsilon_i) = \int_0^\infty f(u_i, \varepsilon_i)du_i$$

$$f(\varepsilon_i) = \frac{2}{\sqrt{2\pi\sigma}}[1 - \Phi(\frac{-\varepsilon_i\lambda}{\sigma})]\exp(\frac{-\varepsilon_i^2}{2\sigma^2})$$

(35)

Where: $\sigma^2 = \sigma_u^2 + \sigma_v^2$. The likelihood function for the sample L can then be derived as the result of the product of all observation-specific density functions, assuming that random errors ε_i for individual observations are independently distributed, hence:

$$L(Sample) = \prod_{i=1}^{N} f(\varepsilon_i)$$

(36)

Taking the log of the likelihood function defined above produces the Log-Likelihood function for the entire sample data.

Aigner, Lovell and Schmidt (1977) define the likelihood function in terms of two variances: the variance of the composite error term $\sigma^2 \equiv \sigma_u^2 + \sigma_v^2$ and lambda λ ($\lambda \equiv \sigma_u / \sigma_v$) which represents the proportion of the inefficiency term's standard deviation σ_u to the random term's standard deviation σ_v which can take any non-negative value. Further to this, Battese and Corra (1977) suggest to re-parameterize λ so that it contains variances instead of standard deviation ($\gamma \equiv \sigma_u^2 / \sigma^2$)[30] such that it becomes bounded between 0 (meaning that all deviations from the frontier are ascribed to noise) and 1 (indicating that deviations from the frontier are entirely attributed to technical inefficiency). Using the λ parameterization, the log-likelihood function of the cost frontier[31] for the sample size with N banks can then be expressed as (Coelli, 1998, p 188):

$$\ln L = -\tfrac{N}{2}\ln(\pi/2) - \tfrac{N}{2}\log(\sigma^2) + \sum_i \ln[1 - \Phi(z_i)] - \frac{1}{2\sigma^2}\sum_i \varepsilon_i^2 \quad \textbf{(37)}$$

Where

$i = 1,2,...,N$,

$z_i = \dfrac{\varepsilon_i}{\sigma}\sqrt{\dfrac{\gamma}{1-\gamma}}$,

$\varepsilon_i = \ln y_i - x_i\beta$, and

$\Phi(z)$ is the distribution function of the standard normal random variable z.

The parameters β, σ^2 and γ are estimated by the MLE technique which finds the maximum of the log-likelihood function in (37) using the observed values of the parameters. This yie/lds estimates of the frontier parameters and the corresponding components of the error term (inefficiency and the random component).

---

[30] The parameter gamma γ is not equal to the ratio of the variance of technical inefficiency effects to total residual variance, that is because the variance of bank-specific inefficiency u_i is equal to $[(\pi - 2)/\pi]\sigma_u^2$ and not just σ_u^2, therefore the actual relative contribution of the inefficiency effect to the variance of the total composite error term γ^* is equal to $\gamma^* = \gamma / [\gamma + (1-\gamma)\pi / (\pi - 2)]$ (Coelli et al, 1998, p 188).

[31] Coelli (1996, p7) states that the log-likelihood function of the production function is identical to that of the cost frontier with some differences in the sings involved according to Schmidt and Lovell (1979).

3.2.8.1 The Maximum Likelihood Estimation Procedure

Parameter estimation according to MLE is conducted using different algorithms, most commonly, Davidon-Fletcher-Powell DFP and Newton-Raphson NR. The DFP requires the vector of first partial derivatives to be derived only, whereas the NR demands the matrix of second partial derivatives to be derived as well, hence it is more computationally demanding (Coelli, 1996). The statistical package[32] utilized in this research to estimate the profit and cost frontiers uses the NR algorithm due to its superiority to the DFP algorithm[33]; that is because the former is more conservative in relation to the convergence of the log-likelihood function.

Under the NR algorithm, convergence is declared when the log of tolerance is equal to 0.0000001 or $1e-7$ (Gould and Sribney, 1999), whereas the log of tolerance is set to $1e-5$ under the DFP procedure (Coelli, 1996). This tolerance criterion is expressed as follows:

$$\left| \frac{l_j - l_{j-1}}{l_{j-1}} \right| \leq l(tolerance)$$

(38)

Where l_j is the log-likelihood function for the set of estimates produced in the j^{th} iteration and l_{j-1} is the log-likelihood function for the set of produced estimates in the previous iteration. That is, the iterative procedure ceases if the proportional change in the log-likelihood function for estimates is less than ($1e-7$). Tolerance specifies the cut-off point where iterations cease declaring convergence of the log-likelihood function and consequently the maximum estimates for the log-likelihood function are produced. More restricted tolerance is believed to deliver more trustworthy estimates since the corresponding value of the log-likelihood will be achieved at a higher possible maximum.

[32] The software used is Stata version 9.

[33] DFP or DFP Quasi-Newton is applied in FRONTIER 4.1. It is acknowledged that there are alternative ML estimation algorithms such as Bendt-Hall-Hall-Hausman BHHH and Broyden-Fletcher-Godfarb-Shanno BFGS, however the statistical package utilized allows for Newton-Raphson algorithm only NR, possibly due to its relative superiority in terms of it requiring the second-order partial derivatives to be calculated besides its relatively conservative convergence criteria.

The maximizing iterative process continues until the log-likelihood function converges – i.e. becomes concave or marginally declining as will graphically be shown in the corresponding empirical chapters. It is important to achieve concavity as the likelihood function may be ill-behaved or convex (i.e. flat) due to having many ridges, flat areas, and saddle points which prevent the function from becoming concave. In such case, more iterations should be allowed so that convergence is eventually attained.

Technically, the NR algorithm for conducting the MLE estimation requires obtaining the first and second partial derivatives of the log-likelihood function (Gould and Sribney, 1999). It follows an iterative maximization procedure to find the closest estimations for the unknown parameters. To initiate this procedure, NR algorithm needs an initial guess – or a starting value – for the parameter estimates to begin the iteration. A gird search is conducted across the space of parameter γ – where $\gamma \equiv \sigma_u^2/(\sigma_v^2+\sigma_u^2)$ – to provide starting values which are then used to initiate the iterative procedure.

Further, if the produced starting values enable the log-likelihood function to be evaluated, the iterative process starts, but if the log-likelihood function is not successfully evaluated using the given stating values, the grid search uses a random number generator to provide a new direction vector until the log-likelihood function can be evaluated, which then triggers the maximizing iterative process. Iterations continue until the log-likelihood function for the set of estimates reaches the maximum value possible. As iterations progress, change in the value of the log-likelihood function – i.e. log of tolerance – becomes smaller and smaller until the process terminates when the difference between the log-likelihoods becomes equal to or less than the logged tolerance as specified in (38).

The procedure that the Newton-Raphson algorithm NR follows to find the maximum likelihood estimates which has a number of steps, as illustrated by Gould and Sribney (1999, p 11). To find the estimate of a parameter b such that the log-likelihood function of which $f(b)$ is maximized, the following steps are taken:

1. The process starts with a guess (starting value) b_0. If the log-likelihood function can be evaluated at b_0, then the maximizing iterative process initiates. Otherwise, the grid search is conducted again to reach an appropriate starting value. A new guess is calculated $b_1 = b_0 - f'(b_0)/f''(b_0)$. This step is repeated and new guesses continue to be generated until the log-likelihood function can be evaluated. The gradient of the likelihood function $f'(b)$ should not be constant (i.e. flat gradient) because this causes the first derivative $f''(b)$ = 0 hence making it impossible to produce new guess according to step 2. This means that the likelihood cannot be increasing or decreasing, suggesting that b_0 is a poor guess then a different starting value is produces.

2. Finding a starting value triggers the iterative maximization process which continues until the log-likelihood function achieves a maximum.

This process of finding the maximum log-likelihood estimate of one parameter b can be extended to accommodate a vector $\mathbf{b}$. Thus, to find vector $\mathbf{b}$ such that $f(\mathbf{b})$ is maximized,

1- The NR process starts with guess $\mathbf{b}_0$.
2- A new guess is calculated $\mathbf{b}_1 = \mathbf{b}_0 + \mathbf{g}(-\mathbf{H}^{-1})$ where $\mathbf{g}$ is the gradient vector, $\mathbf{H}$ is the matrix of second derivatives, and $\mathbf{g}(-\mathbf{H}^{-1})$ is the direction vector. The second step is repeated until the function can be evaluated.
3- The iterative process starts until the log-likelihood function reaches the maximum possible.

Having discussed the technicalities of the log-likelihood function and the maximum likelihood estimation procedure along with the specification of the cost and profit functions that will be used to estimate technical efficiency, the focus now shifts towards discussing the methodological underpinnings of the analysis.

CHAPTER 4: RISK

4.1 Introduction

This chapter explores banking risks including credit, market, liquidity, and insolvency risks from industry, regulatory, and research perspectives. The aim is to provide a background on the underpinning concepts of these risks from different perspectives as they will be used in the following empirical chapters to modify the profit and cost functions. This risk-modification process constitutes the main contribution of this research to efficiency and analyses.

Risk is crucial to the banking business which is centred on taking and managing risk. Risk is "an essential ingredient in bank production" (Hughes et al, 2001, p 2171), and as "the world of banking has changed, risk management must keep pace and research in the area of risk analysis is becoming increasingly important" (The Banker, Sept. 2007, p 17). The global credit crunch which was triggered by problems in the US subprime mortgage market in August of 2007 is a prime example of how important and far-reaching the implications of banking risks can be.

As a result of the credit crunch, which began with the drying up of wholesale credit markets, financial institutions have experienced periods of major falls in the value of assets combined with shrinking market liquidity due to vanishing investor confidence. The spread of the crisis has been the direct consequence of securitization which allows banks to pass on their credit exposures to investors by "slicing and dicing" credit risk exposures using collateralized mortgage obligations (CMOs[34]) for housing loans, and collateralized debt obligations (CDOs[35]) for non-housing loans (Woods et al, 2009).

Under the current crisis, different business models have caused some banks to fail,

[34] A CMO is a special purpose entity which is the legal owner of a pool of mortgages. The entity issues tranches of bonds which are secured against the mortgage pool. Buyers of the bonds include banks, hedge funds and other institutional investors, but the value of the bonds is dependent upon the security provided by the underlying mortgage collateral. If there is uncertainty about the mortgage repayments then the bonds are likely to fall in value very quickly" (Woods et al, 2009, p 32).

[35] "A CDO, like a CMO is a special purpose entity that issues debt against credit-based collateral, but instead of specializing in mortgage debt, CDOs combine various types of debt. A CDO splits the different types of debt into tranches which represent different levels of risk ranging from AAA rated through to the lowest grade known as 'toxic waste'. The tranches are then used as collateral against a bond issue which pays interest in accordance with the level of collateral risk. Higher grade tranches therefore earn lower returns than the junior lower grade tranches" (Woods et al, 2009, p 32).

whilst others have survived, albeit with seriously damaged profits. Northern Rock, which was eventually nationalized by the British government in late 2008, relied on a business model that assumed away any major liquidity crunch in wholesale credit markets, as the Rock's model assumed permanently liquid and deep money markets to support its operations. Other banks around the globe have incurred huge write-downs in asset values such as HBOS which reported, after it was taken over by Lloyds TSB forming Lloyds TSB Banking Group, a total loss of £10bn losses in bad loans, mainly caused by bad loans in the corporate sector (The FT, Feb 13th 2009).

This has forced many big institutions to simply rely on taxpayer funds for survival such as Fannie Mae and Freddie Mac, the main mortgage providers in the US, and Northern Rock and Bradford & Bingley in the UK (Woods et al, 2009). By the end of the financial year of 2008, the five largest UK banks (HSBC, Royal Bank of Scotland, Lloyds TSB, Barclays and HBOS) had recorded write downs and asset impairments totaling approximately to £22 billion (Smith, 2008 and Woods et al, 2009). HSBC, however, reported a 28% fall in profit in the first six months of 2008. HBOS, the UK bank most exposed to the crunch, recorded loan write-offs and trading debt write downs around £2.4 billion (Woods et al, 2009), which later turned out to be over 4 times these losses in Feb of 2009.

In the end, the crisis has shown that increasing credit risk exposure can lead to liquidity problems (the dry-up in wholesale market) which in turn are reflected by increasing market risk (falling share prices) and ultimately increasing risk of insolvency (resulting in forced sales, takeovers and possible bankruptcies). Consequently, this suggests that these four banking risks are crucial to the banking industry and therefore have to be considered upon analysing the performance of banking institutions. The crisis has also shown that banking risks are not only idiosyncratic, but also significantly linked to the overall stability of the financial system, i.e. systematic risk. This implies that risks at bank- as well as banking system-level should be accommodated for in any bank performance assessment process, so as to more comprehensively reflect the nature of the banking business and to increase the reliability of bank performance analysis.

The abovementioned facts constitute a genuine motive to incorporate banking risks into efficiency and analysis. This has been asserted by a number of previous studies.

For efficiency studies, Mester (1996, p 1026) finds that "unless quality and risk are accounted for, one might easily miscalculate a bank's level of inefficiency". Further, Drake et al (2006, p 1451) confirm the vitality of incorporating risk factors into the cost function as "failure to adequately account for risk can have a significant impact on relative efficiency scores". What is more, past efficiency studies seem to completely overlook the impact of regulatory capital on profit and cost efficiencies. This research attempts to bridge this gulf by accounting for Basel I's criteria for measuring credit risk in efficiency analysis. To this end, the credit risk appetite (CRA) measure, calculated as the ratio of risk weighted assets to total assets, is applied. Although Basel I has been implemented for more than two decades, little attention has been paid by efficiency research to accommodating for the impact of Basel I regulatory capital requirements.

As for analysis, Girardone et al (2004, p 225) confirms that "risk and asset quality factors appear to matter in relation to estimation and … these factors should be borne in mind in any future evaluation of the efficiency characteristics of European banking markets". Moreover, Altunbas et al (2000, p 1605) assert that "if risk and quality factors are not taken into account, optimal bank size tends to be overestimated". Accordingly, the contribution of this research to banking efficiency and literatures lies in more comprehensibly accounting for risk by incorporating credit, market, liquidity, and insolvency risks at bank- as well as country-level into the analyses, since past research seem to ignore the impact of some of these major risks.

The reminder of this chapter is organized as follows. Section 2 provides an overview on the forces that are changing the environment of risks which the banking industry is facing. Section 3 discusses the four major banking risks that are later applied in the empirical analysis from an industry and regulatory perspectives. Section 4 examines these risks in academic research and exposes this research's envisaged contribution to the literature. Section 5 is the chapter's summary.

4.2 The Changing Environment of Banking Risks

“Change in the financial environment will stimulate a search by financial institutions for innovations that are likely to be profitable” (Mishkin, 2004, p 232). Such a statement simply highlights the fact that banking risks are changing in nature which is a direct result of the ever-evolving banking business, and therefore research needs to keep pace with these changes, which is what this research endeavours to achieve. It is quite evident that over the last few decades, the banking industry has been experiencing markedly significant changes that have directly influenced their cost base and profitability and has therefore re-shaped the risk-return relationship. As competition from banking and non-banking institutions intensifies, the traditional source of banking income represented by comfortable interest margins has been under considerable pressure.

Banks seem to respond to these challenges in different ways, some of which include: (1) increasing the size of their loan portfolio to make up for the reduced margins, hence running higher loan-to-deposit ratios in their balance sheets such as Northern Rock which had a ratio of over 300%, i.e. for every pound in deposits there are 3 pounds invested in lending (The FT, Sept 17th, 2007), compared to an average EU-8 ratio of around 111% (ECB, 2008a, p 57), (2) taking on more credit risk by allowing for riskier loans to constitute their lending portfolios, an issue which was manifested in the increased exposure to the Subprime market in the US such as in the case of the Swiss bank UBS. UBS had a total exposure to the US subprime market of $27.6bn reported in its 4th quarter statement with half of this exposure turning to real losses which totalled to about $13.7bn (The FT, Feb 14th, 2009), (3) expanding in size to diversify their income sources and to take advantage of cost cuts form the exploitation of. Recent examples include the takeover by RBS of the Dutch bank ABN Amro in Oct 2007, and the recent formation of the Lloyds TSB Banking Group in which Lloyd TSB took over the troubled HBOS in Jan 2009 (The FT, Feb 1st 2009, and the FT, Jan 20th 2009). Such responses by banks have had profound implications on the risk-return trade-off that research needs to account for in some form[36].

[36] Although these are recent examples, they are highlighted to reflect on the changes that the banking industry has been undergoing, which has direct implications on changing the nature of risks and risk management in banking. This will be elaborated on more the coming discussion in this chapter. It is also worth noting that this research does not assess the impact of mergers and takeovers as it is merely concerned with the impact of risk on efficiency and analyses.

To elaborate, although higher loan proportions of the banks' assets should theoretically generate higher long-term earnings, this cannot be decoupled from the greater potential for loan losses due to bad luck (i.e. negative economic conditions) or bad management (i.e. skimping on risk management resources which results in mounting loan losses etc.). At the extreme, a rapidly-growing loan portfolio that is not supported by a proportionate and adequate capital is a sign that the bank can have problems in the long term. This is because counting on positive market movements and having, on that basis, a considerable volume of low-quality loans in the bank's loan portfolio can ultimately translate into solvency problems. The excessive exposure to credit risk in conjunction with inadequate capital structure to support this expansion may push the bank into serious liquidity problems, as has been the case for Northern Rock, one of the largest mortgage banks in the UK. Northern Rock adopted a vulnerable business model which relied heavily on short-term money market funding to support its assets. However, when markets turned sour due to the credit crunch which started in the US in August 2007, a big hole emerged in Northern Rocks' books due to sudden liquidity shortage (short-term funding) which eventually forced the regulator to nationalize the Rock at the expense of the tax-payers' funds (The FT, Feb 17th 2008).

To provide a more detailed view on the forces driving change in the banking business that are related to traditional banking and trading books, the following list, which is may be inconclusive, provides a greater insight into these factors (Goddard et al, 2007), including:

1. Globalization and Financial Crises
2. Technological Advances
3. Securitization
4. Interest Rate Volatility
5. Financial Regulations: the evolution of. Basel II

The following discussion reflects on each of these points and is coupled with real examples from the European banking industry. The ultimate aim of this discussion is to provide a rationale as to why main banking risks (credit, trading, liquidity, and insolvency) introduced in this research should be accounted for upon analysing bank performance (that is in terms of efficiency and analyses).

4.2.1 Globalization and Financial Crises

Globalization is one major factor driving change in banking and therefore posing new risks for banks. Globalization entails that boundaries between countries become less effective leading to greater competition amongst markets and institutions globally. In the broadest sense, globalization could mean that the economic policies of one country can affect that of another country. Therefore, due to globalization and heightened competition, banks found that old ways of doing business are no longer efficient and largely not profitable. Banks have realized that to stay in business, access funds efficiently, and make profits, products and services has to be tailored to meet customers' evolving needs and to prove profitable after all.

For European banking, globalization is of a dual impact. The first dimension of such impact is driven by the integration of the EU economies and the subsequent endeavours to achieve fully-integrated financial markets. The second is the competition arising internationally. The rapid expansion of banks' off-balance sheet activities (OBS) in terms of expansion in debt securitization and derivative trading – mainly due to narrowing margins on the banking book activities – is another profound example of the impact that globalization has on banking. Goddard et al (2007, p 1918) point out that "deregulation and the integration of the European banking sector are likely to have profound implications for the size and concentration of banking markets".

More recent evidence of how globalization has re-shaped the risk-return relationship for European banks is the knock-on effect of the U.S. subprime financial crisis which has wiped out substantial volumes of profits for many European banks: UBS for instance reported a whopping $18bn Subprime-related write-downs in early 2008 (The FT, Feb 31st 2008), BNP Paribas revealed €898m in write-downs in the last quarter of 2007 (The FT, Feb 20th 2008a), and Credit Suisse also reported $2.85bn mark-down (i.e. write-downs) in the value of their assets (The FT, Feb 19th 2008). Many banks sought fresh capital injection to mitigate the profound impact of the cross-Atlantic losses due to mortgage market exposures.

On the positive side, however, globalization provides an avenue for banks to spread their risks across the globe, hence further exploit the diversification effect to reach more preferable risk-return payoff (McLaney, 2006). In this context, investing in diversified or low-correlated business sectors – which is better realized in globalized banking sector – would create less risky portfolio overall due to reduced returns' volatility, yet with greater expected average returns. This in turn can boost both cost efficiency – since lower portfolio risk can lower the cost of funding which ultimately entails less capital to be set aside, and profit efficiency as well – because average return on a diversified portfolio can be less volatile than a risky portfolio which can translate in lower losses overall.

Therefore, it can be argued that globalization is a double-edged sword. On the one hand, it has opened up new avenues for European banks to diversify and expand their business lines, such as HSBC bank in the UK. HSBC has a considerable presence in China and other rapidly-growing Asian markets, a strategy that has smoothen the impact of the current economic slowdown in the US and Europe to an extent. On the other hand, globalization has made one major bank's problems spill over globally – a good example is the increased volatility which global stock markets experienced in March of 2008 that was driven by bad news on the fall out of the 5th largest US investment bank, Bear Sterns, which a considerable number of banks have invested in (The FT, March 17th 2008). The wider implications of such unfolding bad news in the banking sector across the Atlantic has been a notable credit squeeze in the UK mortgage market for instance, with first-time buyers finding it considerably difficult to get a foot on the property ladder due to dramatically fewer mortgage deals being on offer (The FT, Feb 27th 2008). The globalization of the banking business therefore entails the need for accounting for banking risks not only on a micro- or bank-level, but also to account for them on a macro- or banking system-level when analysing banks performance. This is therefore what this research attempts to achieve.

4.2.2 Financial Innovation and Technological Advances

There are three main aspects in which financial innovation that have significantly impacted the management of banking risks (The Economist, May 2007). First is the introduction of credit derivatives (CDs). CDs are used by buyers to pass on credit exposure to investors. They are tradable instruments designed by banks and sold to risk-seeking investors. Second is the expanding role of rating agencies which in many aspects have facilitated the trading of credit risk globally.

Banks increasing reliance on rating agencies in assessing borrowers' credit worthiness has proven to be a risky business. That is because borrowers' ratings have mostly ignored the scenario in that serious liquidity risk can lead to default under the assumption that extreme events related to liquidity risk are very unlikely, an issue that has proven to be a considerable underestimation of reality (The FT, Oct 27th 2007). Despite that banks may have saved costs by outsourcing the task of credit risk assessment, yet, the latest subprime mortgage crisis has clearly demonstrated that excessive reliance on credit rating agencies can be hazardous and sometimes misleading in terms of forecasting wide-scale default events.

The third implication is related to technological advances in risk management. "Thanks to technological and financial wizardry, loans are now made with little contact between borrower and lender, and are shuffled around the financial system like so many cards at a poker table" (The Economist, May 2007, p 42). The ability of taking debt off the balance sheet and spread risks across the financial system has made it easier for banks to further expand in lending and be more complacent about risky loans since they did not have to increase their capital cushions as a result, thanks to securitization. This has had significant implications on risk as one major bank's risky investments can create the potential for a systematic risk. The main driver of this trend has been the considerable investment in technological advances motivated by cost-reduction in information production and management, and the distribution of financial services via the use of electronically-based processes to substitute for the more time- and space-demanding paperwork.

This is evident as banks can nowadays sell on credit risk within few days of generating the asset or even hours. Although this may have contributed to cutting operational costs, however in many cases, this seems to have affected the quality of loan books due to less responsible lending decisions made. As a result, credit has become available to a wide sector of corporates and individuals, even to those with relatively low prospects of success or credit worthiness, since risk could easily and quickly be passed on to investors around the world. This has been one major factor driving a significant and rapid growth in European economies as in the UK for the last 10 years or so. However, this has eventually materialized in mounting billions of risky loans with inherently considerable default risk in the run up to the subprime credit crisis lately.

Besides cost savings, financial innovations and technological advances have also contributed to creating new sources of income for banks, however this technologically-oriented banking business model has been coupled with new risks such as: the risk of identity theft, internet fraud, money laundering etc. It is worth noting here that this shift towards more technology-based banking does not eliminate or reduce typical banking risks such as credit or liquidity risks, however it can change the way these risks are analysed and assessed (Bank Systems and Technology, 2008).

Some banks have chosen to exploit advances in IT and established the so-called 'internet-only' or 'virtual' banks (Mishkin, 2004 and Bank Systems and Technology, 2008). It is a bank with no physical existence, but rather a completely internet-based entity. Egg plc – the UK based virtual bank owned by an insurance company, Prudential – is one example (which in turn is part of the panel data used in this research). Obviously, the cost base, mainly overheads, is much lower for these banks than those of a conventional high-street bank which may give virtual financial institutions a cost efficiency advantage. Risk-wise, standard banking risks still exist for virtual banks; therefore, it still is essential to account for these risks when analysing the performance of these banks.

Looking ahead, Mishkin (2004) argues that in the future, neither the "click" nor "brick" banking business model will dominate the other entirely as large section of customers would still prefer the convenience of face-to-face contacts, let alone their security concerns over obtaining totally internet-based financial products and services. The 'brick' banking has not completely ignored going virtual, but on the contrary, considerable investments are deployed in IT systems supporting their internet banking business, hence offering customers the convenience of using the internet, yet maintaining their branching presence simultaneously.

Advances in technology and financial innovations can increase banks' operational efficiency in providing products and services but albeit with risks remaining present (MacDonald and Koch, 2006). Trading in CDs market is one example. UBS, for instance, seems to have rushed into structured credit instruments trading in the aim of swiftly amplifying profits, but it overlooked rigorous risk management to support the large positions UBS has taken in these markets. This is a typical banking problem as, in good times, thorough risk management practices tend to be overlooked and it is only when problems start emerging banks seem to realize the full scale of these problems (The FT, Nov 2007).

Some analysts have concluded –given the substantial losses and write-downs many European banks have suffered due to the Subprime crisis– that banks need to "go back to basics" and focus more on what they do best, that is, facilitating the payment system and mediating funds between lenders and borrowers, as this shift towards taking excessive risks and being heavily involved in trading seems to becoming problematic (The FT, Nov. 2007). However, banks would still need to take risks to create opportunities and stimulate economic activities. Yet the associated risk should be comprehensively approached and adequately managed in the sense of setting clear risk limits in accordance with the bank's capital make-up.

All the same, it seems that the motivation for banks for investing in IT developments and financial innovations is to enhance their operational efficiency (via cutting costs and boosting profits) has not eliminated typical banking risks, but it has changed the way in which these risks are assessed and managed. This consequently necessitates taking banking risks (credit, liquidity, trading, and insolvency) into account upon analysing banks performance. Part of the impact of financial innovation is accounted for by including Off-balance Sheet items (OBS) as a third output in the analysis given the importance of these items to the bank's income and their impact on the structure of the bank's balance sheet. Technological advances are also assessed by using a panel data set in this research which allows for detecting dynamic changes in inputs and outputs to influence efficiency estimates.

4.2.3 Securitization

Banks have been innovative in responding to the increasing cost of running higher loan-to-deposit ratios in two ways: first, by broadening their fee-based services, and second, by taking these capital-intensive assets off their balance sheets giving securitization a world-wide scale (MacDonald and Koch, 2006). Securitization generically involves bundling a group of homogenous assets and float securities (bonds) on the back of these assets (The Economist, Mar 6th, 2008). In essence, any cash flow-generating asset can be securitized.

Such advances have allowed some European banks – such as Deutsche Bank for instance – to become a financial-liquidity factory through which loans are transformed into marketable securities with considerable profits being made in the process. This has had substantial implications on banks' lending portfolios in terms of greater risk diversification and liquidity. Spreading credit risk globally via securitization takes debt owned by the bank off its balance sheet and therefore releases more funds to be used in other investments, yet, the total risk exposure – although largely spread out – remains systematically present.

In this sense, ignoring the bank's involvement in securitization in analysing bank performance can be undermining. Therefore, this research accounts for off-balance sheet items (OBS) as a third output in both and efficiency analyses, and also uses the ratio of OBS / Total Assets as a determinant variable explaining profit and cost inefficiencies of European banks in an attempt to gauge the impact of the degree of involvement in OBS activities on technical efficiency.

4.2.4 Interest Rate Volatility

The considerable increase in interest rate volatility has been a significant force driving change in demand for banking products in the last few decades. This has put pressure on interest margins which are banks' traditional source of income. To give an idea on how volatile interest rates have been, the following provides some recent examples from the US, Europe, and the UK. In the US, the Fed funds' rate dropped from 5.5% in the 3rd quarter of 2007 down to 3% by early 2008. This was followed by 17 consecutive increases of a 0.25 percentage points from 1% by the end of 2004 up to 5.5% in 2006 and 2007 only to drop again to 3% in January of 2008 after two consecutive cuts within the last two weeks of January 2008 (The Federal Reserve Bank, 2008). The Fed rates experienced a series of cuts in 2008 as a result of the financial crisis reaching a record low of 0.25% in Dec 2008 (The FT, Dec 16th, 2008). In the EU, interest rates tend to be less volatile. The Main Refinancing Operations rate MRO was 3% in early 1999, peaked at 4.25% in June of 2000 and kept falling to reach a trough of 2% in June 2003, only to start to rise again to 4% in 2007 (European Central Bank, 2008). Rates have also recently been cut down to 2% in early 2009 (The FT, Feb 5th, 2009).

As for the UK, the Bank of England's rate starting at 5% in Jan 2007 and rising up to 5.75% until the end of 2007, then rates came back to the 5% level for the most of 2008 but then experienced consequent falls down to a 315-years record low of 1% in Feb 2009 (The FT, Feb 5th, 2009).

It is well established that the trend of interest rates volatility has been less dramatic for banks operating in the Euro-zone (ECB, 2004a), but the fact that large banks are no longer limited to domestic or EU markets remains as banks now have their assets' portfolio spread globally. This makes them prone to be affected by other major economies' interest rates volatility, as that of the US. Under this increased exposure to interest rate risk (volatility), banks have responded by innovating financial products that would mitigate, spread, and transfer their exposures to interest rate risk. Prominent example of this include the adjustable-rate mortgages and financial derivatives introduced in the 1970s which in turn have been under substantial re-engineering ever since – driven by the market's demand and varying conditions. Therefore, accounting for OBS activities – which derivatives trading constitute a significant portion of – seems imperative upon analysing banks performance as this accommodates, to a significant extent, the evolving structure of the banking business. As indicated earlier, this research accounts for OBS items as a third output, and for the ratio of OBS/Total Assets as a determinant factor explaining profit and cost inefficiencies, and it also accounts for the proportion of net-interest income to total income to assess the impact of income structure on European banks operational efficiency.

4.2.5 Financial Regulation: The Evolution of Basel II

Notwithstanding the merits of Basel I in attempting to set international standards for minimum capital requirements aiming to make banks safer financial institutions, Basel I largely focuses on addressing credit risk thus overlooking other major and quantifiable banking risks as in market, liquidity, and insolvency (BIS, 2003, and Rose and Hudgins, 2008). Basel I does not consider the issue of securitization which has increasingly been a major vehicle for banks to transfer risk and a source of fee income. Another loophole of Basel I is that it applies an exogenous "one-size-fits-all" risk weights that have long been criticized for their inflexibility in reflecting banks' exposure to credit risk.

In the 1990s, regulators became aware of the significant impact that market risk exposure could have on the banks' earnings and net worth (capital). This resulted in a modification of the original Basel I of 1988 in 1996 by allowing banks to apply more internally-oriented market risk measures, namely the value at risk or VaR, and allocate capital against the calculated VaRs accordingly.

It is argued that the inflexibility of Basel I's rules has encouraged banks taking advantage of its loopholes and allowed them to run riskier portfolios, which has been counterproductive to the original objective of Basel I (BIS, 2003 and Rose and Hudgins 2008). Securitization is one global financial development that Basel I fail to adequacy assign capital for due to the rather simplistic asset classification it applies in which, for instance, corporate loans and credit card loans were grouped in the same risk-weighted category. As such, banks can shift their portfolios in this particular class to constitute riskier and potentially higher-yielding assets yet incurring the same level of capital charge for that particular category.

Basel II provides a more flexible framework and advocates more internally-developed risk quantification approaches to credit as well as market risks. In fact, Basel II overcomes the one size fits all risk weighting methodology of Basel I, and more closely aligns banks' risk profile to capital requirements. That is why some banks found that in many cases less capital can be set aside to meet the New Accord's requirements (The Economist, May 2007). In 2004, the Basel Committee of the Bank of International Settlements BIS agreed on the modified version of the original Accord, Basel II, to be applied to the largest banks internationally (BIS, 2008). Basel II is due to take effect by 2009 so that to ensure a gradual adjustment of risk management procedures to accommodate Basel II's regulatory requirements.

Basel II constitutes three pillars: **Pillar 1** is concerned with the **Minimum Capital Requirements** that are based upon supervised in-house estimation models for market, credit and operational risk exposures. **Pillar 2** involves the **Supervisory Review** of the banks' risk assessment producers and capital adequacy to ensure their adequacy and reasonableness. **Pillar 3** advocates **Greater Public Exposure** of banks' financial situations in the aim of promoting better market discipline that acts as an exogenous disciplinary factor for risky banks; so as to demonstrate that adequate risk management practices are in place or otherwise to lower their excessive exposure accordingly.

Initially, Basel II will be limited to large banks (over $1bn in total assets), and ultimately all banks applying Basel I will be implementing Basel II. Rose and Hudgins (2008) explain that big banks would be at an advantage in employing their sophisticated risk-quantification models to meet Basel II's requirements which are considerably more risk sensitive, which could yield fewer capital requirements than those under Basel I. This is brought about by the fact that Basel II is structured to ensure consistency and capital-sensitivity to different levels of risk, in the sense that low-risk assets would be associated with less required capital compared to that of risky assets. This in effect would place banks applying Basel I at a disadvantage due to possibly running higher capital ratios. However, reaching a conclusive view on whether Basel II would have such impact on different bank sizes requires a dedicated empirical research, an area that is beyond the scope of this thesis.

Research-wise, notwithstanding the drawbacks of Basel I in adopting the "one size fits all" approach to minimum capital requirements which may have encouraged banks to hold riskier portfolios, attempting to account for the impact of capital requirements on bank efficiency, i.e. the inclusion of Credit Risk Appetite measure, is necessary, as to the best of the researcher's knowledge, no pervious study has attempted to gauge the impact to of Basel I's credit risk weighting on European banking profit and costs efficiencies. It is also worth noting that, from a research perspective, testing the impact of regulatory capital assigned to credit risk can only be approached on the basis of Basel I. This is simply because Basel II has not been fully applied yet, besides the fact that it relies on in-house measures for credit risk, thus it is rather difficult to obtain data on such models for the purpose of this research.

The following section focuses on the major banking risks including: credit, trading, liquidity, and insolvency risks and discusses each risk from a practical and research perspectives. The section concludes by reflecting on how these risks are going to be accounted for in the empirical analysis, and also by highlighting the contribution of this research to both efficiency and literatures from a risk perspective.

4.3 Major Banking Risks

In this section, the four main risks related to the banking and trading books – liquidity, credit, trading, and insolvency risks – are defined and explored from an industry and regulatory perspectives. These risks are then explored in Section 9 from a research perspective in the sense of how they are accounted for to modify the cost and profit models used in efficiency and analyses.

4.3.1 Liquidity Risk

4.3.1.1 Liquidity Risk: Definition

Liquidity risk is associated with the bank's funding needs and the management of its balance sheet's cash flow position. Liquidity exposure arises when the bank falls short of necessary liquidity to fund its due liabilities and its day to day operations. This exposure increases when such funds cannot be acquired at the expected terms and cost when required (HSBC plc, 2006). Liquidity risk, therefore, materializes from holding insufficient liquid assets to fully and timely meet liability commitments when they fall due, mainly those with short-term maturities, forcing the bank to acquire funds more expensively. In principal, banks manage their liquidity positions with the aim of maximizing access to liquidity and minimizing funding costs (Rose and Hudgins, 2008).

4.3.1.2 Liquidity Risk: An Industry Perspective

Banks very rarely run out of cash to meet immediate liabilities such as deposits withdrawals because they hold a variety of liquid assets on their balance sheets, e.g. government bonds that have deep markets. This is in addition to their ability to access money and inter-bank markets – in normal conditions. In practice, banks manage their liquidity positions in the aim of fulfilling all current or short-term deposit withdrawals and other immediate on- and off- balance sheet funding commitments when they mature at the appropriate cost. This also necessitates that access to capital and money markets should be available and should also be cost effective.

Moreover, central banks and financial regulators aim for the wider financial stability and therefore tend to disallow the spread of a systematic bank run at all costs. Thus, in times of crises, the first immediate risk that worries bankers is liquidity and how to manage it in a timely and low-cost manner.

In any case, both occasional and persistent liquidity shortages have undeniably significant adverse impact on the cost of funding for the bank, let alone the negative effect this can have on the bank’s credit rating which directly affects its money markets accessibility. Sound liquidity management process involves maintaining a diversified, stable, and growing core deposit base that is made up of retail and corporate customer deposits. Such basic liquidity tier should also be supported by a second funding tier in terms of ensuring access to wholesale money markets, plus a portfolio of liquid assets which is better being diversified in currency and maturity (Bank of England, 2005).

Most crucially, banks can make provisions for potential liquidity shortages by adequately – but not extensively – investing in liquid assets with deep markets and low liquidating costs. Liquid assets comprise of items that should have deep markets in the sense that turning them into cash can be conducted in a timely and cost-efficient fashion. Liquid assets involve: cash in vaults, cash due from deposits held at domestic banks and Eurodollar deposits, T-bills, government bonds (such as UK Gilts and US Fed funds), Certificates of Deposits CDs, and Commercial Papers CPs.

These lines of defence against liquidity shortages were not in place for Northern Rock, a major UK mortgage lending bank. Northern Rock had a vulnerable funding model: it relied heavily on wholesale markets to raise funds with a rather small core deposit base. When credit dry-up loomed in August 2007, the Rock failed to raise funds even from those banks that used to lend it, and also failed to securitize its loans since confidence in markets was evaporating due to considerable global uncertainty. Eventually the Rock's funding problems were quickly translated into a liquidity crisis that was deepened by the subsequent bank-run on the bank (The Economist, March 3rd 2008).

To avoid such scenarios, managing liquidity risk can follow different practices, including (HSBC plc, 2006): (1) cash flow projections in major currencies which should be related to the level of liquid assets needed to meet the net positions over a given period, (2) daily assessment of liquidity ratios for the bank's balance sheet against the bank's internal limits and regulatory requirements, (3) ensuring the diversity, accessibility, and cost effectiveness of the back-up funding facilities, (4) monitoring the fluctuations of core deposits base and the concentration of which so as to avoid excessive dependence on large depositors, and ensuring a weak correlation between the different funding sources, (5) stress-testing liquidity positions and establish funding contingency plans accordingly, and finally and probably most importantly (6) holding adequate and proportionate amount of liquid assets as sub-optimal levels of liquidity is cost-ineffective, and excessive liquidity tie-up also causes profit deficiency as it reduces revenues.

On the measurement of liquidity risk, every bank may approach liquidity risk differently, however the following examples simply provide an idea on how banks do this in practice in order to show that the liquidity measure this research applies stems from what the banking industry actually applies. JP Morgan Chase for instance uses different approaches to this end: The **Basic Surplus** measure assesses the bank's ability to sustain a 90-days stress event within which no fresh funding can be obtained to fulfil obligations as they fall due (JP Morgan Chase & Co., 2006). Royal Bank of Scotland applies, for example, the ratio of **Net Surplus of marketable assets to Wholesale Liabiliti**es due within one month[37] (RBS Group, 2006). Other ratios include: Cash and government securities / total assets, Purchased funds / Total assets (Rose and Hudgins, 2006, p 215). HSBC plc (2006) gauges its liquidity risk by using the ratio of **Liquid assets / Customer funding** which is very similar to the measure of liquidity risk applied in this research, that is: **Customer and Short-Term Funding / Liquid Assets**. As it can be seen, the liquidity risk measure used in this research is similar to that of HSBC's and perhaps more comprehensive as it considers both types of short-term liabilities, customer deposits and money market funding, as sources of liquidity risk.

4.3.2 Credit Risk

4.3.2.1 Credit Risk: Definition

Credit risk measures the exposure to losses arising from the failure of borrowers (retail and corporate) to meet their debt obligations as they fall due. In other terms, credit risk arises as financial claims associated with the bank's assets may not be paid in full (Saunders and Cornett, 2003). Because banks borrow short and lend long, i.e. hold little proportion of their capital as liquid assets while transforming the majority of their liabilities' maturities into long-term assets, it only takes a considerable proportion of the bank's loans portfolio to default to "push the bank to the brink of failure" as Rose and Hudgins (2008, p 177) put it. Such negative impact of loan defaults is amplified should the loan portfolio be less diversified and more concentrated in a given sector or region.

Therefore, credit risk is one significant risk to take account of upon analysing bank performance. So much so that, in fact, high growth rate driven by rapid expansion in

[37] Net Marketable Assets are: Debt securities + Treasure and other eligible assets + Reverse Repo agreements with banks and customers – [Repos with banks and customers + Short positions "securities to be sold" + Insurance companies' debt securities held + Debt securities charged as security for liabilities]. Net Wholesale liabilities due within 1 month include: Deposits by banks (excluding Repos) + Debt securities in issue – loans and advances to banks (gross, excluding reverse Repos) (RBS Group, 2006, p91)

loans portfolio is normally indicative of potentially forthcoming financial difficulties for the bank, this is because rapid growth is normally associated with taking on lower-quality loans, a trend that has been largely motivated by the spread and ease of debt securitization. The massive losses that the Swiss bank UBS had accumulated in early 2008 as a direct result of excessive exposure to the US Subprime mortgage-linked securities, and the rapid growth of Northern Rock from having around 3% up to a third of the entire UK mortgage market in the last 8 or so years only are prime examples of how serious that mismanaged credit risk exposures could be (The FT, Jan 31st 2008). Thus, accounting for lending-related credit risk in the analyses of bank performance is vital to reaching more reliable estimates. The following section exposes this research's contribution to the literature in accounting for credit risk from an ex ante perspective.

4.3.2.2 Credit Risk Appetite CRA

Part of this research contribution stems from the incorporation of a credit risk appetite measure (CRA) into efficiency analysis. The idea of CRA is drawn from the Bank of England's segmentation of the traditional Return on Equity measure (RoE). This segmentation of RoE is guided by the different sources of banks **profitability** as well as **financial strength**.

The decomposed version of RoE yields the following relationship (Bank of England, 2003, p 74):

RoE = (Pre-tax Profit[38]/ Operating Income) x (Operating Income / RWA) x
(RWA / Assets) x (Assets / Equity)

(39)

These four components denote the following measures respectively: Pre-tax Profit Margin, Risk-Adjusted Asset turnover, Asset-Risk Ratio, and Financial Leverage. This break down assists in identifying which elements of the RoE ratio contribute to profitability or financial stability. According to the Bank of England's Financial Stability Review (2003, p 74), the following table demonstrates the main components of a typical Return on Equity RoE ratio:

[38] Pre-tax profits = Operational Profits – Loan Loss Provisions – Costs.

Ratio	What it measures?	Effect of rise on Financial Stability
Pre-tax Profit Margin	Impact of costs and bad-debts	positive
Risk-Adjusted Asset turnover	Efficiency on a risk adjusted basis	positive
Asset-Risk ratio	**Credit risk appetite**	**Negative**[39]
Financial Leverage[40]	Gearing	Negative

It is evident that RoE can be misleading if it is not decomposed to its main components. For instance, taking the effect of Leverage ratio as an example, RoE will be higher for banks that are highly geared as opposed to their less leveraged peers, suggesting that banks' profitability can improve as the level of its debt increases relative to its total assets or for holding lower equity base. From a financial stability perspective, gearing up can have a direct and positive impact on bank profits up to a certain level, as excessive levels of gearing can threaten the bank's financial stability which in turn can have wider systematic implications. Recent leverage figures in terms of percentage of debt to equity for big UK banks for instance is an evidence of this: "Here are two frightening statistics: over the past five years, the balance sheets of many of the world's largest banks more than doubled; and, according to the Bank, the median ratio of debt to equity in big UK banks is more than 30 to one" which effectively suggests about 3.3% equity coverage for the debt held on their balance sheets only (The FT, Jan 23rd, 2009). In fact, one of the main reasons of the current financial crisis is that banks seem to be undercapitalized given their excessive levels of gearing (The Economist Jan 22nd, 2009). The RoE measure in its entirety, however, cannot provide this perspective on financial stability in relation to gearing. Likewise, an increase in RoE can be driven by a higher Asset-Risk or Credit Risk Appetite ratio which is the result of more credit risk been taken. Although higher RoE in this case suggests higher profitability, it can also suggest a weakening financial strength given the increasing

[39] i.e. higher asset-risk ratio indicates more credit risk taken by the bank.
[40] This can also be referred to as Equity Multiplier (Assets / Equity) which is the inverse of the Capital Ratio (Equity / Assets). More accurately, gearing is calculated as the ratio of debt, net of liquid assets, to the market value of capital (Bank of England, 2008).

exposure to credit risk. Therefore, the suggested decomposition of the RoE ratio is important.

4.3.3 Credit Risk: A Regulatory Perspective

4.3.3.1 The Concept of Risk Weighted Assets RWA

The concept of Risk Weighted Assets RWA was introduced by the Bank of International Settlements BIS as part of Basel I Accord in 1988 to be applied as an international regulatory framework. The aim was to establish a levelled international platform for banks to operate according to and to provide greater safety to financial systems. Basel I proposes a minimum capital ratio that banks should hold of 8% [41](with a Core Capital ratio to be 4% at least)[42].

Briefly speaking, the value of RWA is constructed by classifying the bank's on- and off-balance sheet items into risk categories. Certain credit conversion criteria are applied to off-balance sheets items so as to obtain the credit equivalent of which, hence rearranging them as on-balance sheet elements. Then the appropriate risk weight is applied to each category with the aggregate of the resulting risk weights becoming the bank's RWA (Bank of England, 2002). Basel I of 1988 is the first accord to set international standards for bank capital requirements related to credit risk exposures. Such capital adequacy framework aims at encouraging banks to align credit risk exposures more closely with capital levels.

Basle I suggest two main Tiers which constitute the banks' regulatory capital requirements, these are (Frost, 2007 and Rose and Hudgins, 2008):

Tier 1 or Core Capital which mainly includes: Equity and equity reserves and any other source of long-term finance that can be characterized as equity finance[43].

---

[41] The Standard Ratio should be at 8% = [(Core Capital + Supplementary Capital) / Risk Weighted Assets] or equally [(Tier 1 + Tier 2 Capital) / Risk Weighted Assets], with Tier 1 Capital standing for: Equity + Reserves (cumulated profits), and Tier 2 Capital comprising of: General Provisions, Medium- and Long-Term Subordinated debt (Basel Capital Accord, July 1988).
[42] That is; Core Capital Ratio = (Tier1) / RWA.
[43] This also encompasses: undistributed returned earnings, non-cumulative perpetual preferred stocks, and hybrid equity – which are dept finance usually perpetual instruments that are contractually convertible to equities in the event of bank's default- and minority interests (i.e. shareholders in the bank's subsidiary who have non-controlling

Tier 2 Capital is divided into upper and lower levels. (A) Upper Tier 2 Capital includes: (1) Hybrid debt – e.g. accumulative preference shares and convertible bonds, (2) Revaluation reserves – comprising of investment and property revaluations[44], and (3) General provisions – these are allowed towards Tier 2 Capital up to 1.25% of the Risk Weighted Assets' volume. (B) Lower Tier 2 Capital comprises mainly of subordinated debt[45].

Banks are required to hold a minimum of 4% of core capital ratio, that is, Tier 1 Capital / RWA, and are required to keep 8% of the ratio of (Tier 1 + Tier 2) / RWA, with the amount of Tier 2 Capital being limited up to 100% of Tier 1.

To decide on the level of their regulatory capital, bank's assets need to be classified into "categories" or "buckets" and weighted with specific risk factors that eventually yield category-specific risk weighted assets. The latter are then summed to provide the total of the bank's risk weighted assets. Basel I Accord accommodates for off-balance sheet items (OBS) by returning OBS items to their on-balance sheet equivalents via applying certain conversion factors.

For on-balance sheet assets, risk weights are assigned as follows:

- Cash, reserve deposits, and government bonds bear 0%.
- Deposits with banks of the OECD (Organization of Developed Industrial Economies) bear 20% risk weight.
- Housing loans are associated with 50% risk factor.
- Loans (customer and corporate) and all other assets bear 100%.

As for OBS items, these are mainly characterized as contingent claims that can

ownership of the equity bundle. These are normally shown in the consolidated accounts). Basle, I mandate that Goodwill is to be subtracted (which is a quantified business reputation regarded as assets when acquired by the bank).

[44] These reserves may have the impact of a double-edge sword. Banks benefit from revaluations in good economic conditions, that is; when stock markets are bullish for instance, yet bearish markets and declining property prices put pressures on banks to allocate additional resources to meet the minimum requirements. Regulators only allow for 45% of these reserves to qualify for Tier 2 Capital.

[45] Not all subordinated debt accounts towards Lower Tier 2 level if its maturity falls below 5 years. 20% discount rate applies for each year of maturity that is below the 5 years threshold. Hence, 80% of subordinated debt accounts as Tier 2 capital if it matures in 4 years, 60% of which does so if it matures in 3 years and so forth.

materialize as real asset or liability. As such, they are converted into their equivalent of on-balance sheet items according to certain conversion criteria where conversion factors range from 0% to 100%. For instance, the conversion factor 0% is assigned for Standby Letters of Credit of less than 1 year in maturity that are cancellable at the bank's desire before funds can actually be drawn against the credit facility. Moreover, Long-term Credit Commitments made to private corporations have 50% credit-conversion factor; whereas 100% conversion factor applies to forward purchases, Reversed REPO[46] arrangements, and standard Standby Letters of Credit 'SLC' facility backing corporate borrowings (Frost, 2007 and Rose and Hudgins 2008). Once OBS items are converted into their credit equivalents, bank assets are then categorized into different credit risk-weighted classes as shown in the table below. The table below lists some of the main items found on the bank's balance sheet which is by no means an exhaustive list (Rose and Hudgins, 2008, p 486):

0% Risk-weighted Category • Cash • Government bonds (e.g. US T-bills, UK Gilts)
20% Risk-weighted Category • Cash items in process of collection • Deposits at domestic banks • ABS 'asset-backed securities' • Credit equivalent of SLCs
50% Risk-weighted Category • Secured mortgages
100% Risk-weighted Category • Consumer and commercial loans • Credit-equivalent of long-term credit commitments

[46] Reversed Repurchased Operations (R-Repos), which is the equivalent of over-night lending.

Recent developments have been concentrated on aligning Basel I more closely to credit risk exposures of derivatives trading including futures, options, interest rates caps, floors and swaps, currency swaps and positions in commodities. The exposure associated with these contracts mainly stems from the so-called counterparty risk, that is, the possibility of a bank's customer failing to pay or fulfil the contractual obligations when it matures – leaving the bank exposed to finding a replacement contract, mostly at higher costs (Rose and Hudgins, 2008). This is the case for OTC 'over the counter' or tailored derivative contracts. However, counterparty risk diminishes if derivatives are traded through an Exchange market such as the London International Financial Futures Exchange LIFFE, since the latter guaranties the fulfilment of derivative contracts should a customer ffailto do so, which entails that banks would not normally be required to set capital against such standardized trading arrangement.

Basel I converts OTC contracts to their credit-equivalent amounts and treats them consequently as on-balance sheet items, which are then multiplied by the appropriate risk weight that will eventually constitute part of the entire balance sheet's RWA. For instance, interest rate derivatives with maturity of 1 year and less have '0' credit-conversion weight, while a 0.5% factor applies to those contracts with more than 1 year of maturity. On the other hand, Basel I associate higher credit-equivalent weighting to currency-based contracts. For example, currency-related derivatives with 1 year or less and those with more than 1 year to maturity bear a credit-conversion weight of 1% and 5% respectively. All the same, this research accounts for credit equivalent of OBS items as a third output in both efficiency and analyses.

4.3.3.2 The Standardized Approach SA

Basel II offers a regulatory perspective to credit risk measurement via the Standardized Approach (SA). The aim of discussing the SA is to shed further light on the way credit risk will be approached under Basel II. This approach attempts to overcome the "one-size-fits-all" criteria of Basel I and is more risk-sensitive in the sense that it more closely aligns risk weights with asset classes. The SA specifies five credit grades corresponding to the level of each asset class's exposure. Credit grades are further defined according to Standard & Poor's S&P ratings which specifies a range of credit rates for each of the 5 credit grades as: AAA to AA-, A+ to A-, BBB+ to

BBB-, BB+ to B-, and below B-. For instance, residential mortgages have 35% risk weight, and corporate loans with credit grades 1 and 2 receive 20% and 50% risk weights respectively[47].

As for securitization, Basel II bridges the regulatory gap of Basel I by allocating capital against this activity. Securities created are split into two trenches (Frost, 2007): (1) Credit-enhanced securities, the value of which is deducted from the regulatory capital requirements (½ from Tier 1 and ½ from Tier 2), and (2) Non-credit enhanced securities issues held by the bank as an originator or an investor are assigned risk-weights on the basis of external ratings. With regard to Off-balance Sheet items OBS, they are converted into their risk-weighted asset equivalents according to their maturity and recoverability using the following conversion factors (Frost, 2007, p 507 – 508): unconditional commitments that are cancellable at any time and those up to 1 year and over 1-year maturities are assigned 0%, 20% and 50% factors respectively. Short-term letters of credits have 20%. Securities pledged as collateral and those lent by the bank have 100% risk weight. Obviously, the application of SA from a research perspective is rather difficult given that it requires detailed knowledge of the asset types and credit ratings attached and so forth.

4.3.3.3 Credit risk: A Practical Perspective

The management of credit risk involves managing the exposure both before (ex ante) and after (ex post) taking the lending decision. This entails examining borrowers' credit profile or credit worthiness before granting loans in order to minimize the adverse selection effect, and to invest in monitoring loans over the life of the credit facility so as to minimize the moral hazard effect. This research takes both perspectives to credit risk and applies an ex ante as well as an ex post measures of credit risk by employing the Credit Risk Appetite and Loan Loss Provisions ratio respectively.

There are few concepts to define before discussing the measurement of lending-

[47] Moreover, assets with grades 3 and 4 are assigned 100%, whereas grade 5 assets have 15%, and all unrated corporate loans have 100% risk weight. For unsecured past-due loans, if with NPL cover < 20%, have 150% weight assigned for all grades, yet if the NPL cover is between 20% - 50% then a 100% risk weight is applied, and for NPL cover > 50%, the risk weight ranges from 50% -100%. NPL cover is the Non-Performing Loans cover = specific provisions/ exposure – eligible CRMS (Credit Risk Mitigation Items, such as collateral, guarantees, and credit derivatives)

related credit risk. **Loan loss provisions LLP** or the so-called "allowance for bad debt losses" are reserves set aside to hedge against potential loan losses. The allocation of LLPs is a subjective decision that can be used to smooth earnings (Frost, 2004). **Non-performing loans** comprise of loans that are past due for 90-days or more. **Charge-offs** are loans that are acknowledged as worthless hence written off the bank's lending book. Any possible income from these technically written-off loans could be generated reduces the level of charge-offs yielding **Net Charge-offs** (Rose and Hudgins, 2008).

It is worth noting that LLP differs from the loan loss reserves in that the former represents a given year's expected credit losses, whereas the latter is the aggregate of current and past LLPs. It is also worth noting that the LLP stands for an ex post assessment of loans credit risk since LLPs are allocated after lending decisions have been made (Frost, 2004, p 78). Therefore, this research applies the **LLP** ratio that is largely applied by commercial banks which is calculated as **LLP / Total Loans** to provide an ex post perspective to the bank's on-balance sheet lending-related credit risk (HSBC plc, 2006). The ex post nature of provisions for loan losses stems from the methodology according to which provisions are allocated. It assumes that impairment losses have occurred but yet not realized as an actual expense at the balance sheet's date. Therefore, provisions are calculated to reflect the value of non-performing or potentially-problematic loans including those that have past-due for 90-days, however, there remains the possibility that all or part of these impaired loans could be repaid in the future, and thus actual losses on these loans are individually identified at some point in the future (HSBC plc, 2006).

4.3.4 Market Risk

4.2.4.1 Market Risk: Definition

Market risk is "the exposure to an adverse change in the market value of portfolios and financial instruments caused by a change in market prices or rates" (JPMorgan Chase and Co., 2006, p 77). This suggests that market risk comprises of two types of risks related to trading and banking (or non-trading) books. The next sections address these two risks following Hempel and Simonson (1999), Heffernan (2002), Saunders and Cornett (2003), Allen et al (2004), JP Morgan Chase & Co (2006), and RBS Group (2006). Greater emphasis will be put on the market trading risk as it will be incorporated in the empirical analysis.

4.2.4.2 Non-trading Risk

Non-trading risks are associated with the banking book which embodies the commercial banking business line. The main sources of market non-trading risk are interest rate risk, equity risk, currency risk and basis risk. Interest rate risk arises fundamentally from asset re-pricing and maturity mismatches between interest-sensitive assets (ISA) and interest-sensitive liabilities (ISL) over a given period of time. This applies to on- and off-balance sheet items.

There are two main sources for maturity mismatches: (1) specific changes in interest rates that are related to a specific asset or liability, and (2) systematic changes as represented by in the yield curve (or the so-called yield-effect) which embodies changes in the outlook of the relationship between long- and short- term interest rates (the yield-effect is particularly important for banks as they borrow short and lend long).

Moreover, non-trading equity risk is associated with equity investments the bank holds in other institutions that mainly include, for example, a venture capital portfolio and other direct equity investments held in other institutions. Other financial instruments constituting the non-trading investment portfolios include debt securities, equity shares, deposits, certificate of deposits or CDs …etc (RBS Group, 2006).

As for the non-trading currency risk, such exposure stems from any foreign exchange (FX) holdings, debt, or foreign investments the bank has. This is related to the so-called Basis risk, which is the risk associated with narrowing spreads between lending and borrowing rates (Saunders and Cornett, 2003). In FX for instance, basis risk arises from the difference between forward and actual spot rates on maturity. For government bonds for instance, basis risk represents the difference between cash (immediate-delivery) and future (postponed-delivery) market prices, such as the basis risk associated with a 10-years maturity UK Gilt that is worth today £95 for £100 face value but is worth £87 for 6-months delivery, thus for example, the basis risk in this case is

8% [= (£95 - £87)/100]. This margin would narrow, i.e. basis risk will increase, if interest rates were to increase, therefore future prices will discount more heavily down to, say, £80. Consequently, this amplifies the basis risk from 8% up to 10%. With regard to the banking book, basis risk arises when, say, mortgage rates and rates paid on deposits move disproportionately hence exposing the bank's interest rate spread to the potential of downside basis risk, i.e. narrowing spreads.

In the context of this research, non-trading market risks were not integrated into empirical models because their measurement requires access to in-house data, whereas this research has access to publicly available data only. For instance, for non-trading interest rate risk the most common measures of which are GAP analysis and Duration analysis techniques. Both models require some detailed in-house data on ISAs and ISLs, besides the need for clear assumptions about how these interest-sensitive items should be classified according to different maturity buckets.

With respect to currency risk exposures related to foreign investments, again, it is a challenging task to quantify such risk as banks' treasuries usually use self-hedging techniques such as international netting and the management their net currency exposure positions by matching cash inflows and outflows of similar maturities. Therefore, most of the non-trading market risks are self-hedged anyway. Another obstacle is data accessibility. So, for instance, detailed in-house data on the management of non-trading currency risk is a major difficulty. Finally, for non-trading basis and equity risks, it is also difficult to quantify both risks given the detailed data required on ISA and ISL to gauge the overall magnitude of the basis risk. Likewise, detailed information on the net position exposure for each of the banks' foreign equity investment portfolios requires access to internal data.

For all the above-mentioned reasons, the non-trading part of market risk was excluded from the risk-modification process of the banking efficiency and models. The focus therefore is on trading risk which is the subject of the next section.

4.2.4.3 Trading risk

Trading risk arises from the potential that adverse market conditions (rates and prices) can negatively affect the level of net trading income (Deutsche Bank, 2006). There are three main sources constituting the trading-book market risk, including: interest rates trading instruments, credit spreads[48], and foreign exchange.

Trading essentially involves taking short (to sell) and long (to buy) positions in equity, debt securities, foreign exchange, commodities, and derivative (futures, forwards, options and swaps) markets. The aim of taking trading positions and assuming trading risks is to profit from spreads between bid (buying) and offer (selling) prices of the underlying asset. Also, banks can make trading profits from undertaking 'arbitrage' – that is, taking on offsetting positions in various yet related markets so as to make profits by exploiting markets' imperfections (RBS Group, 2006). The following section provides an idea on how trading risk is measured in practice.

4.2.4.4 Trading Risk Measurement: A Practical Perspective

Banks apply various statistical and non-statistical (accounting-based) approaches in quantifying trading-related market risks. Statistically-oriented measures mainly involve: Stress testing (scenario analysis) and Value at Risk (VaR). The following sections will elaborate on these two techniques and then focuses on how trading risk is accounted for by this research.

4.2.4.4.1 Stress Testing Model

Stress testing is concerned with examining the 'what if' events and their potential impact on trading risk using both historical and hypothetical scenarios to this end (Société Générale Group, 2007). Under the historical stress testing, past financial and stock market crisis are analysed over a given period of time –say starting from the Black Monday event of October 17th 1987 up until the ERM crisis in 1992 or through to the Latin American crisis of 1995 and the 1997 Asian crisis…etc. Subsequently, changes in equity prices, interest rates, exchange rates, and credit spreads are analysed during these crises. The ultimate aim is to establish several 'potential

[48] Spreads are the risk premiums associated with a given security: that is the difference between risk-free and risky instrument such as equities and commodities.

scenarios' that can then be applied to the bank trading positions and ultimately estimate the level of possible losses under such stress scenarios.

For hypothetical Stress Tests, the bank designs the so-called 'what-if' extreme yet plausible hypothetical scenarios involving potential sequence of events that, if realized, could result in a financial crisis –as in a fall of 20% or 30% of all stock market indices which the bank invests in– due to political unrests, economic downturns in a major economy…etc. Combining the two stress testing methodologies would produce very useful information upon which loss limits and trading exposure limits can be established, hence avoiding any "open-ended" exposures should the assumed adverse market conditions materialize. Finally, it is probably worth noting in this context that in the UK, the Financial Services Authority (FSA) conducts an independent worse-case or stress testing assessments for UK banks which lately seemed to be accurate in projecting forthcoming difficulties in the banking sector.

Lord Turner, the head of the FSA commenting on the currently revealed HBOS's losses of around £10bn, said recently that these losses were not unexpected according to the FSA's stress test conducted in Oct 2008, given HBOS's considerable level of 'potentially' bad loans on its balance sheet (The FT, Feb 25th, 2009).

4.2.4.4.2 Value at Risk (VaR)

VaR, by definition, is "the minimum likely loss over the next trading day" that a bank can incur on its trading position under a given confidence limit (mostly 99% or 95%) and a given time-horizon (such as 1 day) (Dow, 2002, p 8). Technically, calculating the VaR requires the determination of a time horizon over which the estimates of changes in the value of the market portfolio will be conducted. This is coupled with choosing the suitable confidence interval that sets the limits to the degree of certainty associated with these estimates. Practically, banks normally assume one- or ten-trading day as time horizon and a confidence level of 95% or 99%.

Methodologies employed for the VaR estimation include: (1) *Historical Simulation* which basically assumes that historical patterns apply to future scenarios in relation to the variance and magnitude of the historical returns (Dowd, 2002), (2) the Variance-

Covariance technique imposes normality on the historical returns' distribution (instead of relying on the actual distribution of historical data) and provides estimates of the expected value and the variance on the basis of which. It also assumes that the change in portfolio value is a linear function of the changes of individual assets comprising the portfolio (Allen et al, 2004), and finally (3) the *Monte Carlo* approach which constructs hypothetical future scenarios for a portfolio's returns using the 'Bootstrapping' technique, where outcomes are randomly-generated using historical trading patterns. Bootstrapped values can have a distribution that is closer to normality as the number of trials is increased the (Dowd, 2002).

On the attractions of VaR, Dowd (2002) states three main merits of VaR. The first is that "it provides a common consistent measure of risk across different positions and risk factors" (Dowd, 2002, p 10) suggesting that VaR is a powerful risk measurement tool in the sense that it allows for measuring risks across different asset classes, such as fixed-income and equity positions. The second advantage is that VaR accommodates for the *portfolio diversification effect* or risk-offsetting effect which acknowledges that individual risks are not perfectly correlated such that the overall risk of a diversified portfolio would be lower than the sum of the risks for each asset comprising the portfolio. Thirdly, as the VaR highlights the maximum possible loss over a specific time frame for a given confidence level under normal market conditions, it provides a useful basis for allocating capital in line with the level of exposure to trading risks.

The following table provides a real example on VaR as produced by JP Morgan bank in its 2006 published annual accounts (JP Morgan Chase and Co, 2006, p 78):

	2006			2005			At December 31,	
As of or the year ended December 31, (in millions)	Average VAR	Minimum VAR	Maximum VAR	Average VAR	Minimum VAR	Maximum VAR	**2006**	2005
By risk type:								
Fixed income	**$ 56**	**$ 35**	**$ 94**	$ 67	$ 37	$ 110	**$ 44**	$ 89
Foreign exchange	**22**	**14**	**42**	23	16	32	**27**	19
Equities	**31**	**18**	**50**	34	15	65	**49**	24
Commodities and other	**45**	**22**	**128**	21	7	50	**41**	34
Less: portfolio diversification	**(70)**[c]	**NM**[d]	**NM**[d]	(59)[c]	NM[d]	NM[d]	**(62)**[c]	(63)[c]
Trading VAR	**84**	**55**	**137**	86	53	130	**99**	103

[c] Average and period-end VARs are less than the sum of the VARs of its market risk components, which is due to risk offsets from portfolio diversification. The diversification effect reflects the fact that the risks are not perfectly correlated. The risk of a portfolio of positions is therefore usually less than the sum of the risks of the positions themselves.

[d] Designated as not meaningful (*NM*) because the minimum and maximum way occur on different days for different risk components, and hence it is not meaningful to compute a portfolio diversification effect.

JP Morgan calculates VAR using a one-day time horizon and a 99% confidence level, according to its 2006 annual accounts. This implies that the bank expects to incur losses greater than that predicted by VAR estimates only once in every 100 trading days, or about two to three times a year. In 2006, the table above shows that for fixed income, foreign exchange, equities, and commodities and other trading assets, average trading VaR was $56m, $31m, $22m, and $45m respectively which sums to a total trading VaR of $154m. By taking portfolio diversification into account, which is $70m representing around 45% (= 70/154) of the summed VaRs of the individual components, average trading VaR for the bank was down to $84m. This suggests that the bank is confident that in 99 out of 100 trading days, the maximum expected trading loss will not exceed $84m. Compared to 2005 figures, the bank clearly had more diversified portfolio in 2006 given the greater impact of the portfolio diversification. This was reflected in the larger overall trading VaR for 2005 of $86m.

Having discussed VaR strengths, it is also worth reflecting on its weaknesses. A key shortfall of the VaR methodology is its lack of *sub-additivity*; suggesting that it is not always the case that the VaR of a portfolio or combined positions would be less than the VaR of the individual components combined: “the risk of the sum could be greater than the sum of risks measured by the VaR” (Dowd, 2002, p 12). More of a serious loophole is the normality assumption of the VaR, especially if such underlying assumption is overlooked by the VaR users or severely violated by real losses. The fact that some assets’ returns in the trading book would inevitably have a ‘fat-tail’ probability distribution – i.e. severer losses for the less likely 5% or 1% events – is not integrated in the VaR system (Dowd, 2002). In fact, the normality assumption of VaR can be the single weakness that may have put an end to its life: events unfolding for the past 12 months in the banking sector have been anything but normal, an issue that has considerably undermined the normality assumption of VaR

Furthermore, there is also the important issue of ‘skewed’ as opposed to ‘symmetrically-assumed’ distributions by the VaR. Moreover, the accuracy of the VaR estimations can be questionable. Taleb (1997, p 37) explains that “you’re worse off relying on misleading information than on not having any information at all. If you give a pilot an altimeter that is sometimes defective, he will crash the plane. Give him nothing and he will look out the window". To mitigate for such drawback, banks tend to back-test VaR estimates. Back-testing effectively serves as a reality-check of the original VaR estimates that is conducted using the distribution of the real profits made and losses incurred on a given portfolio over the same time period assumed by the VaR.

The answer to the VaR's insensitivity to 'fat-tail' events is to measure market trading risk using the Expected Tail Loss (ETL) methodology. In essence, ETL is the expected value of the tail-losses, i.e. losses exceeding the VaR value, which can be expressed as: ETL = E [L | L > VaR] (Dowd, 2002, p 29). The marked advantage of ETL over VaR is that ETL is a 'coherent'[49] risk measure, whilst the VaR is not. Consequently, ETL suggests more reliable and accurate estimates that those of the VaR.

The VaR was not applied to efficiency and analyses for two main reasons. First is the issue of consistency in published VaR data. Banks use different methodologies, confidence intervals, and holding periods in calculating their VaRs. Therefore, a considerable degree of inconsistency can be embedded in published VaR values for the sample banks which may ultimately undermine the reliability of estimates. Second, despite that VaR is used in the banking industry as a standard measure for trading risk that is wildly relied on in published accounts, VaR itself suffers a major weakness related to its normality assumption. Therefore, although it might be an interesting research idea to integrate VaRs in the analysis, however it is believed to add little value given the extreme abnormality that unfolding events have implied in the banking sector since late 2007.

---

[49] A coherent risk measure satisfies four conditions according to Dowd (2002, p 28). So, If X and Y are two risky positions, $\rho(.)$ is considered being a coherent risk measure if the following properties are met:

- Sub-additivity: $\rho(x)$ + $\rho(y) \leq$ $\rho(x+y)$
- Homogeneity: $\rho(tx)$ = $t\rho(x)$ where $t > 0$
- Monotonicity: $\rho(x)$ $\geq$ $\rho(y)$ if $x \leq y$
- Risk-free condition: $\rho(x+n)$ = $\rho(x) - n$ where $n > 0$

The second and third conditions imply that $\rho(.)$ is a convex function. The fourth condition means that the addition of a sure amount n to the position will decrease our risk by the same amount as it will increase the value of the portfolio by the end of the period.

As far as the ETL is concerned, obtaining detailed historical returns data on each trading portfolio in the trading books of each of the sample banks involved was, technically, quite implausible given the type of in-house data required. This is in addition to the fact that the distribution of returns for each individual bank's trading portfolios in the sample should be established, this is along with the two properties of skewness (asymmetry) and the level of kurtosis (fat-tail) for each bank in order to reach a creditable ETL estimates. For these reasons, ETL was ignored as a trading risk measure, although calculating ETL and integrating estimates into efficiency and analyses would be an interesting project should the data become available.

3.4.5 Insolvency Risk

3.4.5.1 Insolvency Risk: Definition

Insolvency risk arises when the bank assets are worth less than its liabilities or what it owes (The Economist, Mar 19th 2008). Thus, a bank is considered as technically insolvent when the bank's Net Worth (or equity capital or assets net of liabilities) is less than its debts. This can be the direct result of incurring considerable losses in the banking book, as it has mostly been the case in the current financial crisis, or trading book which can exhaust the bank's net worth (Allen et al, 2004). Although it is rare for banks to go insolvent, however recent bank bail-outs in Europe and the US has shown that insolvency is a real risk that has long been assumed away, and therefore research needs to take account of. In fact, The Economist (Sept 25th, 2008) concluded that banks are no longer immune to the insolvency threat which can only be cured by beefing up their equity capital. This actually meets with what this research has found in relation to the impact of insolvency risk on European banks' operational efficiency as will be shown in the first empirical chapter.

3.4.5.2 Insolvency Risk: Case Studies

Notwithstanding the rarity in which insolvency risks can materialize as bank failures, the aim of this section is to reflect on some real examples on how serious insolvency risk can be at a micro level as well as at a systematic level. A good example from the US banking system is the case of Continental Illinois National Bank which failed in

1984 (Saunders and Cornett, 2003, p151). Continental was expanding aggressively in the energy-based lending market with an annual growth ration of about 20% between 1977 and 1981. Such rapid growth is not always a healthy sign as Continental's capital was not structured proportionately to accommodate the fast expansion of its loan portfolio.

To be more specific, the bank had a narrow core deposit base and heavily relied on wholesale funds, mainly short-term, inter-bank and money markets funds – Fed funds via Repo arrangements – and Eurodollar deposits (dollars deposed with the bank outside the US). However, as the US economy experienced a significant downturn in early 1980s, this contributed to increasing Continental's credit risk exposure hence resulted in substantial defaults on its sizable energy-related loan portfolio. Due to the scale of defaults on the bank's ill-diversified lending portfolio, Continental's credit worthiness and its ability to make full and timely repayments on its short-term liabilities became questionable and led to difficulties in renewing or rolling over its wholesale funding facilities which eventually was suspended; putting Continental at a significant exposure to liquidity shortages. The compounded negative impact of substantial credit and liquidity risks besides its fragile funding model inevitably left Continental technically insolvent. The Fed ultimately claimed control of the bank in 1984.

For the UK banking system, history seems to repeat itself as there seem to be little evidence that lessons have been drown from Continental's failure. Northern Rock, the British bank ranking amongst the top 5 mortgage lenders UK-wide, has experienced similar difficulties to that of Continental's and ended up being nationalized by the UK government in February 17th 2008 (The FT, Feb 17th 2008). Northern Rock adopted a brittle funding structure to support its assets and the rapid rate of growth in its mortgage portfolio. The bank relied excessively on wholesale funds, and took the renewal of such market credit facilities for granted. Unfortunately, when the US Subprime crisis hit, banks almost stopped lending funds to each other because of lack of transparency in terms of the scale of potentially undisclosed losses as a consequence of the crisis. Consequently, the Rock was left to dry up of liquidity over a very short period of time in August/September of 2007. The Rock's insolvency problem was therefore the consequence of poor liquidity, which in turn was a direct result of a fundamentally vulnerable funding model.

Dowd (2007a) argues that the liquidity crisis and the subsequent state of insolvency the Rock eventually fell in was of its own making, and stresses that banks would maintain their own financial health only if they have an incentive to do so. Dowd (2007a) further argues that "financial institutions should live or die under the discipline of the market". However, the majority of central banks governors believe that financial institutions should not be left to fail, especially big ones, given the systematic implications on the wider financial system (The FT, Feb 17th 2008). The arguments on wider systematic implications of one bank failing and whether or not the government should bail-out tumbling banks using taxes is a different issue. Regardless, the core issue here is that the Rock is to blame for not establishing sound risk management practices such as stress-testing its exposure to markets dry-up (Dowd, 2007b), let alone the fundamental imbalance it had between its fast-growing loan portfolio and its fragile capital structure.

The conclusion that can be drawn form the above-discussed examples of well-established banks falling apart and becoming insolvent is that: although financial regulators or central banks cannot afford to allow a bank to fail and become insolvent for the far-reaching systematic ramifications, banks can become insolvent and therefore accounting for insolvency risk in assessing bank performance would certainly deliver more credibility to the analysis results. For that reason, a measure of insolvency risk in terms of capital ratio (equity capital / total assets) is incorporated into efficiency and analyses at both bank-specific and country-specific levels.

3.4.6 The Integration of Banking Risks in Empirical Analysis

The following sections elaborate on how the four main banking risks discussed are integrated into efficiency and analyses of this research. This is followed by a summary section which highlights the contribution of this research to the current literature from a risk perspective.

3.4.6.1 Credit risk: A Research Perspective

Credit risk can be associated with banking as well as trading book exposures. In this research, the focus – due to data limitation – is on credit risk related to the banking book which arises from realizing non-performing or past-due loans. This is accommodated for by the incorporation of an ex post proxy of credit risk in efficiency and analyses both at a country- as well as bank-specific levels. The proxy is represented by the ratio of **Loan Loss Provisions / Total Loans**. This ratio has traditionally been employed by past research as in Mester (1996), Clark (1996), Berger and Mester (1997), Altunbas et al (2000), and Casu and Girardone (2004). However, this research takes a step further and accounts for credit risk from an ex ante (external to the bank) perspective, that is, a measure of credit risk exposure prior to any loan impairments or losses occurring. This is achieved by the introduction of Credit Risk Appetite measure (CRA) suggested by the Bank of England (2003) such that **CRA = Risk Weighted Assets / Total Assets**. CRA is introduced to efficiency analysis only as an explanatory variable to the inefficiency term due to data limitations. To the best of the researcher's knowledge, no previous banking efficiency research has applied this measure to date, bearing in mind that calculating the level of RWA is a regulatory requirement of Basel I Accord which has been in place for about two decades.

3.4.6.2 Market Trading Risk: A research perspective

Banking efficiency studies seem to completely ignore the impact of trading risk on bank performance. This research attempts to bridge this gap by integrating a measure of trading risk at bank- as well as country-specific levels as represented by the ratio of **Net Trading Revenue / Total Revenues**. This ratio is believed to provide information on (1) the level of exposure to trading risk and on (2) the risk-return relationship related to trading. The absolute value of the trading risk ratio is indicative of the level of exposure to trading regardless of whether it is positive (reflecting trading profits) or negative (reflecting trading losses), whereas the sign of the trading risk ratio reflects the risk-return pay-off in trading. So, if net trading revenue is relatively large and positive, this suggests that there is a relatively considerable exposure to trading risks which is paying off because of the positive net trading revenue (trading profits) achieved.

3.4.6.3 Liquidity risk: A Research Perspective

Banking efficiency research seems to largely overlook accounting for liquidity risk which is one of the most serious risks as the current financial crisis has shown. The only study that have applied a proxy for liquidity risk at bank- and country- level in a European context is, to the best of the researcher's knowledge, Altunbas et al (2000). Therefore, to provide more reliable analysis to banking performance, liquidity risk has to be addressed. This research, accordingly, incorporates a proxy for liquidity risk into efficiency and analyses. Liquidity risk is measured as **Customer & Short-term Funding / Liquid Assets**. This ratio is introduced both at country- and bank-levels. This ratio measures the ability of one bank to fulfill its short-term commitments or liabilities using liquid assets. The higher the ratio, the greater the liquidity risk is and vice-a-versa.

Liquidity risk measure is incorporated at a bank-specific level in efficiency analysis as a determinant variable of the profit and cost inefficiency terms (empirical chapter 1), and as a control variable in the cost function used to estimate (empirical chapter 2). To account for liquidity risk at a country-specific level, the liquidity ratio is averaged for the number of banks operating in a given European country following BankScope's (the database utilized) geographical classification. The aim is to control for the impact of banking systems' differences in terms of liquidity risk on the profit and cost efficiencies for the small sample of 541 banks (empirical chapter 1), as well as on the using a bigger sample of 2026 banks (empirical chapter 2).

3.4.6.4 Insolvency risk: A research perspective

This research accounts for insolvency risk both in the efficiency and analyses. Insolvency risk is proxied **by Capital Ratio = Equity Capital/ Total Assets**. In this sense, the lower the ratio, the greater the possibility for the bank to become insolvent i.e. the less stable the bank's net worth is under financial pressures. Previous studies have accounted for insolvency risk in different ways. For instance, Spong et al (1995), in Berger and Mester (1997), Maudos et al (2002), Bos and Kolari (2005) and Kasman and Yildirim (2006) use capital buffer ratio (equity capital / total assets) to proxy for insolvency risk. Altunbas et al (2000), however, use the level of financial capital to proxy for insolvency risk. All past research applies insolvency risk measure on a bank-specific level. This research takes a step further and applies insolvency risk measure at a country-specific level as well in both efficiency and analysis.

Lastly, it is worth noting that accounting for credit, trading, liquidity, and insolvency risks at a country-specific level helps in isolating the exiguous impact that these factors can have on banks' estimated inefficiencies and. This procedure is motivated by the observation of Berger and Mester (1997).

3.4.7 Summary

The table below summarizes the risk variables this research uses to modify bank efficiency and analyses and highlights the gaps in the literature that this research attempts to bridge (as shown in the last column #5 to the right of the table). The table also shows whether the risk variable is applied at country-specific and/or bank-specific levels for each of the two samples and indicates if any previous empirical work has taken the corresponding risk measure into account, hence highlighting this research's envisaged contribution.

Table 1: Risk Measures and their Application

(1) Risk Variable	(2) Description	*(3) Profit / Cost Efficiency Analysis*	*(4) Scale Economies Analysis*	(5) Accounted for by previous research?
Liquidity Risk	Customer & SR Funding ÷ Liquid Assets	*Accounted for at bank- and country-level*	*Accounted for at bank- and country- level*	Accounted for at a bank-level only (Altunbas et al, 2000)
Credit Risk (ex post)	Loan Loss Provisions ÷ Total Loans	*Accounted for at country- and bank-level*	*Accounted for at bank- and country- level*	Accounted for at a bank-level only (Altunbas et al, 2000)
Credit Risk (ex ante)	Risk Weighted Assets ÷ Total Assets	*Accounted for at bank- level only (because the aim is to investigate its impact on operational efficiency)*	*Not accounted for due to data limitation*	**No**
Trading Risk	Net Trading Income ÷ Total Revenue	*Accounted for at country- and bank-level*	*Accounted for at bank- and country- level*	**No**
Insolvency Risk	Equity Capital ÷ Total Assets	*Accounted for at bank- and country-level*	*Accounted for at bank- and country- level*	Accounted for at a bank-level only (past studies mostly used raw equity capital to this end, e.g. Altunbas et al, 2001)

It is probably worth noting that this research utilizes accounting measures for banking risks, as in Altunbas et al (2000, 2001, and 2007), as the majority of European banks do not have publicly traded securities. Risk measures used in this research to modify efficiency analyses may seem quite simple at face value, yet, empirical results have shown the significant impact these measures can have on profit and cost efficiencies as well as on estimates. This is besides the fact that the vast majority of past research has approached risk in a similar manner. The contribution of this research lies in accounting for a wider range of risks more comprehensively. The proposed risk proxies are not only applied at bank-specific level as the vast majority of past research do, but also at a country-specific level to account for country differences and to control for the exogenous impact these risks might have on bank profits or costs (the dependent variables). Despite some statistical limitation to the measures, the inclusion of risk extends the boundaries of efficiency and research significantly as will be demonstrated in the following empirical chapters.

Sophisticated and statistically-oriented risk models may not, after all, be the answer given their underlying assumptions. Such assumptions may, in abnormal circumstances, result in decoupling these models' outcomes from their ability to anticipate real problems. Allan Greenspan – the former Federal Reserve chief – in commenting on the latest financial unrest that has been sparked by the US mortgage crisis reasserted the importance of having sound and capable risk models in place, but equally admitted that "these models do not fully capture what I believe has been, to date, only a peripheral addendum to business-cycle and financial modelling – the innate human responses that result in swings between euphoria and fear repeat themselves generation after generation with little evidence of a learning curve". Greenspan concludes that "we will never have a perfect model of risk" (The FT, March 16th 2008). Notwithstanding this reality, this research suggests that it is informative to recognize the link between the different types of risks and banking efficiency and analysis. The risk variables introduced have demonstrated significant impact in both analyses and were shown to be vital to obtaining accurate and unbiased efficiency and estimates.

CHAPTER 5: EMPIRICAL ANALYSIS

5.1 Research Sample and Data

The research sample comprises of an unbalanced panel dataset of 541 observations on large European banks (with minimum total assets of US$1bn) that were active over 10-year period between 2008 and 2018 (inclusive) and located in 14 European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Norway, Spain, Sweden, Switzerland, and the U.K.

Data was obtained from Bankscope resource package of 2018 produced by Bureau Van Dijk (BVD) which is widely used in the literature [e.g. Altunbas et al (2001), Cavallo and Rossi (2001), Casu, Girardone, and Molyneux (2004), Kasman and Yildirim (2006), Yildirim et al (2007), and Fitzpatrick and McQuinn (2008)]. The sample contains monetary values obtained from balance sheet and profit and loss accounts. The analysis uses data from consolidated accounts only so as to avoid the potential problem of double counting for globally-operating institutions[50].

Moreover, the justification for using a panel data set stems from the argument of Altunbas et al (2000, p 1611) in that “technical efficiency is better studied and modelled with panels”. This is because panel data allows factors with changing effects over time to be modelled and consequently affect the estimated profit and cost efficiencies. This in turn will reduce the potential for biasness in estimates that can result from ignoring these changes.

It is also important to use consolidated data in the analysis (i.e. data produced on an organizational-level rather than on a subsidiary-level). This is because the use of unconsolidated (subsidiary-level) banking data allows for several factors to create some bias in the data. These factors can include: shifting profits between subsidiaries hence affecting the reliability of the estimated profit efficiency results, and sharing inputs or any other intra-organizational arrangements as pointed out by Berger et al (2000) and Bos and Kolari (2005).

[50] Financial statements are consolidated on the basis of country of operation, i.e. the data is produced in Bankscope as country-based consolidated accounts. For instance: ING Bank has many branches across Europe such as ING Bank (France), ING Bank (Netherlands) and ING Bank (Deutschland). Each of these banks is accounted for as a separate bank in the data set used because each is operating in different countries.

Furthermore, the study focuses on large EU banks only (with minimum total assets of US$1bn) so as to minimize potential scale-bias which could result from estimating a common (single) efficient frontier for both small and large European banks (Casu and Girardone, 2004). This is because large and small banks may have different production technologies (i.e. different input mixes and prices and different output levels and mixes) as explained by McAllister and McManus (1993), hence fitting a single frontier for a sample of banks with distinctively different production technologies and cost structures is very likely to produce unreliable estimates. Accordingly, focusing on large banks only helps mitigating this problem as estimating a single frontier for the entire sample which inherently assumes that all banks have similar production technologies (McAllister and McManus, 1993).

Moreover, the study focuses on commercial banks only so as to avoid producing biased efficiency estimates due to different production technologies or other effects assumed by analysing non-commercial banks (Fitzpatrick and McQuinn, 2008). Therefore, the study analyses commercial banks because they have similar production technologies (e.g. output mixes), an assumption that is also consistent with estimating a single frontier for the entire sample (Bos and Kolari, 2005). Although large commercial banks can have different insurance businesses lines and investment banking windows, their core business is related to banking book activities, i.e., their activities are mainly related to taking deposits and lending which qualify them to be classified as commercial banks.

More importantly, Monetary values were deflated using country-specific GDP deflators with the year of 2008 being considered as the base year. Deflators were produced on the basis of Consumer Price Index (CPI) data obtained from same database, Bankscope of 2018. Row CPI data appeared to have different base years for the different countries involved therefore deflators (CPIs) were rescaled to have a unified base year of 2008. So, for instance: if raw data shows that the CPI for 2008 = 0.95 using 2018 as base year (meaning that the CPI for 2018 = 1), then using 2008 as the new base year will rescale the CPI for 2018 to be 1.0526 (= 1/0.95), while CPI for 2008 will become equal to (1) and so forth. The resulting deflators were then applied to the corresponding country-specific monetary values. The deflated monetary values were subsequently deployed in the estimation of profit and cost frontiers.

5.1.2 Standardizing data denominations

Since raw monetary values extracted from Bankscope of 2018 have different currency denominations including Danish Koruna (DKK), Swiss Frank (CHF), UK pound (GBP), euro (EUR) …etc, a Euro-standardization procedure was necessary to neutralize the currency effect and therefore achieve consistency in monetary denominations across the different 14 countries. The process involved obtaining end-of-year euro exchange rates from Bankscope to convert the relevant currency denominations into euros for the entire dataset.

It is important to show how Euro-standardization process was conducted to the different currency denominations of balance sheet items employed in the analysis. Row data extracted from Bankscope of 2018 for the 14 European countries involved in this analysis was denominated in 6 different currencies in including the euro. Banks from 5 countries had their data denominated in domestic currencies. This required that pair exchange rates with the euro to be obtained. Exchange rates were used to return all monetary values expressed in different currency denominations for these 5 countries to euros so as to achieve consistency in data denominations before conducting the estimation. Data for the remaining 9 countries was already expressed in euros because this analysis covers the period between 2008 and 2018 which is after the introduction of the euro in January of 1999.

5.1.3 Variable Specification 1: Frontier Variables

Table 2 below provides definitions and descriptive statistics of outputs, input prices, risk factors, and other control variables over the sample period of 2008 – 2018 for 541 observed banks that are employed to estimate profit and cost frontiers.

Table 2: Frontier Variables

Variable[51]	Description	Mean	Std.	Min.	Max.
P_{it}	This is bank-specific Profits representing the dependent variable in the profit frontier function that is normalized by (equity capital $\times$ price of labour) and transformed such that: P_i **= [Profits + |Minimum Profits| + 1]**	290.4314	91.43972	101.7152	1220.853
TC_{it}	This is bank-specific Total Costs representing the dependent variable in the cost frontier function that is normalized by (equity capital $\times$ price of labour) such that: TC_i **= Total Costs / (equity capital $\times$ price of labour)**	212.2038	649.4181	10.78203	7167.011
w_1	**Price of funds / Price of labour** <u>Where</u>: (1) Price of funds = interest expenses/ (customer deposits + short-term funding + other funding). (2) Price of Labour = Personal expenses / Total assets	6.74572	16.58709	.3079903	143.1213
w_2	**Price of fixed capital / Price of labour** <u>Where</u>*:* Price of fixed capital = (Other admin + Other operating expenses) / Fixed assets	985.0027	1311.287	2.583052	7429.839
y_1	**Loans / Equity Capital** <u>Where</u>*:* Loans = Total Loans + Total Problem Loans – Loan Loss Reserves	11.82536	4.242392	1.451099	23.5694
y_2	**Other earning assets / Equity Capital**	8.890328	6.383754	1.402524	50.58273
y_3	**Off-balance sheet (OBS) items / Equity Capital** *(the credit equivalent of OBS)*	6.537319	6.249379	1.013514	40.92739
t	**Linear form of the time effect**	-	-	1	5

[51] It should be noted that profits, costs, input prices and outputs are presented in de-logged format in as these terms will be logged when forming in the functional form to be estimated. It is a standard practice in the literature to present summary statistics of frontier variables in logs.

				(2008)	(2018)
$\frac{1}{2}t^2$	**Non-linear form of the time effect**	-	-	0.5 (2008)	12.5 (2018)
EU	**EU** membership dummy	-	-	0	1
$EURO$	**Euro** adoption dummy	-	-	0	1
GDP	**Country-specific GDP growth rate**	2.032018	1.548122	-.2742983	9.191397
π_k	**Country-specific Inflation Rate** Where k =1, 2…14 countries.	2.236773	.823539	.2016129	5.524239
r_k	**Country-specific SR interest rates** *(3 months Interbank rates)*	3.160949	1.251038	.32917	6.9075
$M2_k$	**Country-specific financial intermediation proxy** = M2 money supply[52]/GDP *(this is a country-averaged measure)*	10.71031	73.23825	.028205	618.5877
$HERF_k$	**Country-specific Herfindahl Index for loan market concentration** *(this is a country-averaged measure)*	.2453979	.1633645	.0974596	.8044983
$C.TR_k$	**Country-specific Trading Risk** = Net Trading Income / Total Revenue *(this is a country-averaged measure)*	.0382202	.0230007	.0183452	.1260356
$C.CR_k$	**Country-specific Credit Risk** = Loan loss provisions / Total Loans *(this is a country-averaged measure)*	.0067729	.0032692	-.0001864	.0232085
$C.LR_k$	**Country-specific Liquidity Risk** = (Customer and Short-term Funding) / Liquid assets *(this is a country-averaged measure)*	26.33519	7.695861	19.701667	49.18933
$C.InsR_k$	**Country-specific Insolvency Risk** = Equity capital (equity + equity reserves) / Total Assets *(this is a country-averaged measure)*	.0929344	.049738	.0402072	.171592

Frontier variables presented in the first 7 rows in the table above are further defined and explained in the following sections. This will be followed by a brief discussion on descriptive statistics on the remaining macroeconomic and country-specific variables.

[52] M2 measures the monetary supply in terms of the Broad Money which includes the value of cash currency in the system plus banks deposits.

5.1.4 Dependent Variables Specification (Profits & Total Costs)

This section illustrates the steps according to which the dependent variable (P_i) of the profit frontier and (TC_i) of the cost frontier are computed.

The specification of profits requires specifying total revenues first, and second, deriving total net income. Total revenues are defined as:

Total Revenues = Interest Income + Net Fee Income + Net Commission Income + Net Trading Income + Other Operating Income

This leads to defining net total revenues as:

Net Revenues = Total Revenues – Total Costs (incl. Loan Loss Provisions)

Loan loss provisions (LLPs) are considered as costs and therefore should be deducted from total revenues to reach net revenues. This is because LLPs constitute a buffer against bad debt that is set aside as expense and are used to build up loan loss reserves to hedge against 'expected losses' in loan portfolio (Drake et al, 2006).

Net revenues need to be normalized by the product of equity capital and price of labour. This is because the right-hand side variables, input prices and outputs, in the frontier function are normalized by the price of labour and outputs are normalized by equity capital respectively following Berger and Mester (1997). Accordingly, hence the dependent variable, profits in this case, has to be normalized by the product of equity and labour price. This procedure ensures that variables on both sides of the frontier equation are consistently normalized by both the price of labour and by equity capital. Accordingly:

Profits **= Net Revenues/ [Equity Capital $\times$ Price of Labour[53]]**

[53] Price of Labour is proxied by the ratio of Personal Expenses to Total Assets.

(40)

The rationale for normalizing Profits by equity capital and price of labour is to strip out the effects of scale-bias and wage differences across sample banks (Berger and Mester, 1997)[54].

It is worth noting that this research applies a different input price normalization procedure to that of Girardone et al (2004) where the authors normalize input prices by the price of fixed capital rather than the price of labour. This research thoroughly tests for the different input price normalizations and concludes that the most suitable procedure involves normalizing input prices by the price of labour. This will extensively be discussed later in this chapter as part of the process of reaching the preferred functional form to estimate.

Following Berger and Mester (1997), the final step involves producing

$$P_i = [\text{Profits} + |\text{Minimum Profits}| + 1]$$

(41)

As the dependent variable (P_i) will be logged in the functional form of the alternative profit frontier, such transformation ensures positive values of (P_i) for all observations. To illustrate, since some banks might make losses or negative Profits as defined in (40) in a given year(s) over the sample period, therefore applying no transformation to Profits means occasionally taking the log of negative Profits which is impossible. Therefore, the observation with the minimum Profits (or maximum losses) is chosen and the absolute value of which is taken and added back to Profits. This step alone would mean that the observation with the minimum Profits will have a value for Profits of zero which also makes it impossible to log. For this reason, the value of (1) is added

[54] Akhigbe and McNulty (2003) normalize profits by assets, rather than equity and find that small banks have higher profit efficiency. Akhigbe and McNulty (2005, p 290) argue that normalizing by equity produces the opposite result since large banks use more leverage (less equity) than small banks. However normalizing profits by equity capital, as Berger and Mester (1997) explain, has more of an economic rationale as it is closer the ROE ratio which should show greater profit efficiency for larger banks (given their relatively lower equity base). Normalizing by assets is likely to produce opposite results (lower profit efficiency for larger banks compared to small banks as large banks hold more assets). This research follows Berger and Mester (1997)'s approach in normalizing profits by equity capital.

to complete the transformation. This way, the observation with the minimum Profits (or maximum losses) will have the variable Profits as defined in (41) equal to 1 as its corresponding Profits and minimum value of Profits will cancel out. Accordingly, such transformation makes it possible to express Profits for all observations in logs.

Turning to the cost side, equation (42) below defines total costs (TC_i) which is the dependent variable in the cost frontier's functional form. First, total costs are calculated as:

Total Costs = Interest Expenses + Personal Expenses + Other Admin Expenses + Other Operating Expenses

(42)

Total costs are then normalized by the product of equity capital and the price of labour to neutralize scale- and wage-difference effects as illustrated earlier. This yields the dependent variable (TC_i) such that:

TC_i = Total Costs / [Equity Capital $\times$ Price of Labour]

(43)

Total costs variable (TC_i) is then logged to be used as the dependent variable in the cost frontier's functional form.

5.1.5 Inputs and Outputs specification

As illustrated in the methodology chapter, the intermediation approach found by Sealy and Lindley (1977) is used in specifying inputs and outputs of banks. The distinctive feature of this approach is to account for deposits as inputs. Consequently, inputs of the banking production process include **labour, funds** (financial capital)[55], and **real capital** (fixed assets)[56]. Banks are assumed to operate under perfectly competitive

[55] Funds (or financial capital) comprises of equity and equity reserves, deposits and short-term funding, and other long-term finance.
[56] Capital (or fixed assets) includes bank's premises, equipment … etc.

input markets (Bos and Kolari, 2005). The prices of these inputs are defined as follows:

1. Price of **Labour** = $\frac{Personal\ Expenses}{Total\ Assets}$

2. Price of **Funds** = $\frac{Interest\ Expenses}{customer\ deposits + S.R.\ funding + other\ \ funding}$

3. Price of **Real Capital** = $\frac{Other\ Adm'n\ Expenses + Other\ Operating\ Expenses}{Fixed\ \ Assets}$

The rationale for approximating the price of labour by the ratio of personal expenses to total assets is twofold.

First, it was practically difficult to obtain data on employees' numbers and their wages for each bank in this research panel. Second, Maudos et al (2002) explain that the ratio (PE/A) of personal expenses (PE) to total assets (A) can be interpreted as labour cost per worker (PE/L) adjusted for differences in labour (L) productivity (L/A), such that $PE/A = PE/L \times L/A$.

As the prices of funds and real capital are normalized by the price of labour, input prices that are accounted for in the profit and cost frontiers' functional forms are therefore defined in (44) and (45) below. According to Berger and Mester (2003), normalization by input price (price of labour) is applied to achieve the homogeneity restriction[57]. Denoting the normalized price of *funds* and the price of *real capital* as w_1 and w_2 respectively, they will therefore be specified as:

$$w_1 = \frac{price\ of\ \ funds}{price\ of\ labour}$$

(44)

[57] The homogeneity restriction should be applied to the cost function but does not have to be imposed on the alternative profit function, however it is applied to the latter to keep the functional forms equivalent as Berger and Mester (2003) explain.

$$w_2 = \frac{price\,of\ capital}{price\,of\ labour}$$

(45)

On the other hand, the literature typically recognizes three main outputs of the banking production process, these are: **loans, other earning assets, and off-balance sheet items**. These three outputs are believed to cover all outputs generated by the entire bank's balance sheet. As can be noticed, the specification as such caters for a third type of output: off-balance sheet items (OBS). It is argued in a number of studies such as Jagtiani and Khanthavit (1996), Altunbas et al (2000), and Altunbas et al (2001) that OBS should be accounted as an output when modelling bank cost function given their increasing importance as a source of income for banks, otherwise the bank's output would seriously be understated. This indeed is evident as far as this research sample is concerned. It is found (as shown in Table 3) that OBS items account for around 1/3 of European banks total assets on average over the study period 2008 - 2018. Thus, accounting for OBS is imperative to achieving credibility in the analysis. Thus, the three outputs are defined as follows:

1. **Loans** = total loans net of total problem loans and loan loss reserves
2. **Other earning assets (OEA)** = deposits with banks, different types of listed and non-listed securities, T-bills, Bonds, CDs, equity investment…etc.
3. **Off-balance sheet items (OBS)** = credit-equivalent of off-balance sheet assets and liabilities.

As illustrated earlier, these outputs are normalized by equity capital to strip out scale-bias effect. Denoting loans as (y_1), other earning assets OEA as (y_2), and off-balance sheet items OBS as (y_3):

$$y_1 = \frac{Loans}{Equity\,Capital}$$

(46)

$$y_2 = \frac{OEA}{Equity\,Capital}$$

(47)

$$y_3 = \frac{OBS}{Equity\,Capital}$$

(48)

The resulting normalized output variables above, y_1, y_2, and y_3, will be incorporated in the efficient frontier functional form that is to be estimated. The next section reflects on the arguments underpinning the need to incorporate and normalize by equity capital in frontier analysis given its significant impact on estimates as will be demonstrated. This will then be followed by the summary statistics for bank-specific variables that are employed as profit and cost inefficiency term correlates.

5.1.6 Accounting for Equity Capital

It is crucial to include equity capital in the cost and profit functions for many reasons. First, Hughes et al (2001, p 2174) explain that equity capital serves as a source of funding, a cushion to absorb loan losses and financial distress, and a signal of asset quality in terms of the resources allocated to maintain this quality to outside creditors. Hence it is important to account for equity capital in the cost and profit functions. Second, Berger and Mester (1997) argue that the inclusion of equity in the analysis corrects for the potential *scale bias* arising from differences in bank dependence on borrowed funds since large banks tend to rely more heavily on borrowed funds and therefore employ less equity capital (in relative terms) than smaller banks. Third, equity capital is an alternative source of finance to deposits in funding loans and investments hence it would ultimately affect the bank's costs and profits. In this sense, ignoring equity capital may produce *biased inefficiency* estimates in favour of large banks or those relying less on equity capital given that equity is the most expensive source of finance. Fourth, in terms of risk, Mester (1996) argues that, other things held constant, a higher level of equity suggests lower default risk. The potential of default risk is reflected on the bank's cost and profit via the risk premium the bank has to pay for borrowing. It follows that, the inclusion of equity is also believed to capture the risk preferences of the bank as risk-averse banks would tend to hold higher levels of equity – although incurring higher costs – than that of a cost-minimizing bank. Therefore, equity capital has been used in the literature as a proxy for default risk the level of which would ultimately impact the bank's cost and profits. Thus, unless risk is accounted for, efficiency estimates might be misleading due to potential bias[58].

On the treatment of equity capital, some studies have employed bank equity capital as a control variable in the cost function to account for risk. For instance, the study by Altunbas et al (2000) find that financial capital has a significant impact on bank's costs that is greater than that of non-performing loans and liquidity ratios which are used to account for credit and liquidity risks respectively. However, the noticeable impact of financial capital should be treated with caution according to the authors, because its influence appears to stem from the number of its interactive terms with other variables in the cost function, whereas parameters accounting for the other two risks (credit and liquidity) are incorporated with no interaction with other variables. The rationale for accounting for equity capital as a proxy for risk as such is that the level of equity capital is believed to be responsive to the level of financial risks the bank takes such that, with more risks taken, more equity capital is required, and this in turn translates into higher financing costs for the bank since equity is the most expensive source of finance.

The inclusion of equity capital in the cost function as such has been criticized in the literature due to the oversimplification in representing risk. Akhigbe and McNulty (2003, p 312) argue that it is rather simplistic to include equity capital in the cost function "to control, in a very rough fashion, for the potential increased cost of funds due to financial risk". This study avoids the inherent scale-bias stemming from applying raw equity capital as a control variable by introducing equity capital as a ratio to total

[58] Mester (1996, p 1026) adds that it is the level rather than the price of equity capital that should be employed in the cost function. She states that "Financial capital should be accounted for in the cost function and the level rather than the price of financial capital should be included since there is good reason to believe that cost-minimization does not fully explain a bank's capital level – e.g., regulations set minimum capital-to-asset ratios and bank managers may be risk averse" (p 1026).

assets as a variable in the cost function.

More importantly, equity capital is used to normalize dependent and output variables in the frontier model following Berger and Mester (1997 and 2003). Accordingly, rather than using equity capital as a control variable, it is used as a netput to deflate or normalize the dependent variable and output quantities. This way scale-biasness is neutralized to a large extent. Netputs are basically inputs or outputs which are taken to be fixed and/or where input prices are difficult to measure. Berger and Mester (2003) specify off-balance sheet items (OBS), real capital (premises and equipments), and equity capital as netputs. The authors account for OBS items in terms of a 'fixed' netput using Basel I OBS risk weights as a rough proxy since these items are difficult to measure according to the authors. This research, on the other hand, accounts for the credit-equivalent of OBS hence OBS items are included as a third output.

Further, the argument of Berger and Mester for treating real capital as fixed netput is that it is difficult to obtain data on the prices of these inputs. Fixed or real capital is not treated as netput in this research because it is possible to proxy the price of which by taking the ratio of other admin and non-operating expenses to fixed assets as shown in the earlier section. In this research, only equity capital is specified as netput for two reasons: first, because it is difficult to measure the price of equity capital (in terms of the risk-adjusted return on equity) across the 14 European countries involved in the study, and second, because equity capital is difficult to change quickly.

Furthermore, the rationale for using equity capital to deflate profits, costs, and outputs is to neutralize scale-bias as Berger and Mester (1997, p 918) explain that "since the costs and profits of the largest firms are many times larger than those of the smallest firms, large firms undoubtedly would have random errors with much larger variances in the absence of the normalization. In contrast, firms of different sizes have ratios of costs or profits to equity that typically vary only by a few-fold. This is particularly important because the inefficiency term is derived from the composite residual, which might make the variance of the efficiencies dependent on bank size in the absence of normalization". In this sense, normalization is important so as to reduce heterogeneity of banks under analysis. Thus, normalizing by equity capital is also necessary to reduce heterogeneity-bias in efficiency estimates by keeping variables away from being skewed in favor of the larger banks. This is particularly important given the considerable level of heterogeneity (variability) in bank total assets of this research.

The sample data of 541 observations shows a minimum, maximum, and mean bank size of €984.6186mil, €1125494mil, and €190918.8mil respectively with a standard deviation of €240967.5mil. It is clear that the sample banks demonstrate a considerable level of variation given that the standard deviation is about 200 times the size of the minimum size bank in the sample. Such normalization process is therefore believed to level the ground upon which banks' efficiencies are compared to one another because "large banks will tend to have higher profits for a given set of prices, primarily because they were able to gain size over a period of decades, a feature that small banks cannot achieve in the short run. However, the profits per dollar of equity and costs per dollar of equity of large banks are well within the achievable range for small banks.

The next section presents summary statistics for bank-specific variables that are used to explain the variation in the profit and cost inefficiency terms. These variables will specify the inefficiency term functions which will simultaneously be estimated with the corresponding cost and profit frontier functions.

5.1.7 Summary Statistics: A Brief Discussion

This section briefly discusses summary statistics presented in Table 3 (page 203). Summary statistics shows that traditional banking activity (lending) relative to equity capital, represented by (y_1), seems to dominate other types of business lines in European banking given that the average size of performing loan portfolio net of provisions over 2008 – 2018 was about twice the average size of off-balance sheet items relative to equity capital as represented by (y_3). Still, this indicates to the significance of OBS activities as a bank output which some studies fail to account for and therefore underestimate the scale of banking output (as in Vennet, 2002). Further, the size of other earning assets relative to equity capital as shown by (y_2) seems to be the second most important output in European banking.

OBS items are accounted for as a third output using their credit risk equivalent. This is based on the Basle Accord of 1988 credit conversion factors of 100%, 50%, 20%, and 0% which aim at converting OBS items into the credit equivalent amount of on-balance-sheet assets. Taking account of OBS items as an output is very important. Earlier studies confirm the significance of accounting for OBS activities such as Boyd and Gertler (1994) and Clark and Siems (2002) using US data. The argument of Clark and Siems (2002) is that, over the 90s, the banking industry has experienced a dramatic increase in financial derivative activities by more than 600 percent. Accordingly, “noninterest income, which is heavily influenced by OBS activities, has increased as a percentage of total income from 19 percent in the late 1970s to nearly 46 percent by 1999, thus, it is increasingly important to include these activities in any evaluation of efficiency of commercial banks” (p 988).

As for macroeconomic variables, the rationale for introducing the GDP growth factor (*GDP*) is quite important, since banks tend to do well in good times (positive GDP growth) but not so in bad times (sluggish or negative GDP growth). This is recognized in the literature as an exogenous factor impacting the efficiency of banking institutions. Berger and Mester (1997) model this as a ‘bad luck’ factor and find strong evidence supporting its effect. Further, other studies demonstrated that changes in macroeconomic conditions can have significant influence on bank performance. Barros et al (2007) for instance observe that bank profits are pro-cyclical – i.e. positively correlated with business cycle movements – while the level of loan loss provisions is counter-cyclical.

Moreover, Drake et al (2006) demonstrate the importance of accounting for macroeconomic factors when analyzing bank profit and cost efficiencies. This research accommodates for Drake et al’s observation and includes, in addition to the GDP factor**,** three more country-specific macroeconomic variables including: inflation rate (π_k), base interest rate (r_k), and a proxy for the level of financial intermediation ($M2_k$) as in Kasman and Yildirim (2006). The proxy for financial intermediation ($M2_k$) reflects the level at which banks are involved in the economic activity. Therefore, the higher this proxy the greater the degree of financial intermediation in a given economy.

Statistics in Table 2 show a considerable dispersion in ($M2_k$), reflecting distinctive differences amongst the 14 European banking systems involved in this study. Again, this stress the fact that accounting for country differences is vital given the pro-cyclical nature of the banking business. The average of this financial intermediation proxy of 11 suggests that M2, broad money supply (which accounts for bank deposits), is 11 times the value of GDP.

On the other hand, the loan market concentration index ($HERF_k$) reflects a distinctive dispersion in the level of competition amongst the 14 European countries involved. The higher the $HERF_k$ index the less competitive (more concentrated) the loan market is because fewer banks will be dominating a large share of the market. Statistics show a maximum $HERF_k$ of about 80% reflecting a highly concentrated loan market, and a minimum of under 1% reflecting a highly competitive loan market. On average, $HERF_k$ score of about 24% represents a relatively competitive European loan market over the period of 2008 to 2018.

The idea of producing a Herfindahl index for loans ($HERF_{Loans}$) stems from the Loans work of Berger and Mester (2003) and Berger et al (2007) who use this index to gauge deposit market concentration for a number of banks operating in a given country. $HERF_{Loans}$ was produced and incorporated into the efficient frontier functional form instead of $HERF_{Deposits}$ for two reasons. First, to the best of the researcher's knowledge, no previous research has applied Herfindahl methodology to gauge the level of competition (concentration) in the loan market. Herfindahl index seems to have only been applied to deposit markets. Second, convergence of the log-likelihood function could not be achieved if $HERF_{Deposits}$ is specified possibly due to the strong collinearity it showed with other macroeconomic variables. Also, including both $HERF_{Loans}$ and $HERF_{Deposits}$ was also inappropriate due to the very high collinearity between the two indices, therefore only $HERF_{Loans}$ is accounted for.

The EU membership dummy (EU) and euro adoption dummy ($EURO$) were introduced on a country level to investigate the potential impact of operating in different economic environments could have on profit and cost efficiencies of the sample banks. It would be interesting to see if the UK banking system– which is an EU- but non-Euro-economy – has any significant impact on profit or costs of UK-based banks compared to the French banking system for instance– which is an EU- and Euro-economy as well, or would opting out of both the EU and Euro would provide better environment for banks to operate more efficiently as it is the case for the Swiss banking system.

With regard to time dummies, Battese and Coelli (1995) explain that introducing time dummies as stochastic frontier variables serves the purpose of detecting shifts in the frontier due to technological changes over time. This research introduces linear (t) and non-linear ($\frac{1}{2}t^2$) forms of time dummies following Altunbas et al (2007) to capture any possible shifts of the efficient frontier brought about by technological advances. The introduction of time dummies to capture possible technological changes as such was also applied by a number of earlier studies such as Jagtiani and Khanthavit (1996), Altunbas et al (2000), and Huang and Wang (2001).

As for country-specific risk factors, country-specific trading risk factor ($C.TR_k$), which is proxied as country-averaged ratio of net trading revenue to total revenue, shows a mean value of just under 4% over 2010 – 2018[59].

Furthermore, country-specific credit risk factor ($C.CR_k$) shows that the average proportion of 2008 provisions relative to total loans was about 0.7% over the study period between 2008 and 2018. European banks seem to have been able to maintain low $C.CR_k$ ratio possibly because of well-diversified loan portfolios and good quality collateral in general.

With regard to country-specific liquidity risk factor ($C.LR_k$), represented by the ratio of

[59] The size of net trading revenues is believed to be reflective of the level of exposure to trading activities. This is because the greater the absolute value or the size of Net trading revenue reflects the bank's exposure to trading activities, whereas the sign of which represents whether there are losses or profits being made from trading. That is why Net rather than Gross trading revenue is considered because Gross trading income does not provide the extent of exposure to trading as it ignores potential trading costs. This is because unless costs are deducted, it is not possible to make inferences about the risk-return trade-off related to trading activities.

customer and short-term funding to liquid assets, summary statistics show that over 2008 – 2018 European banking systems seem to hold customer deposits and other short-term funding as liabilities that are, on average, about 26 times the value of their liquid assets. In other words, the liquidity cover over that period was about 4% (1/26) on average. This relatively low ratio for liquidity cover (or high $C.LR_k$) indicates increasing reliance on customer deposits and short-term funding resources (numerator effect), or reduced liquid assets holdings (denominator effect). The former view is somehow supported by the findings of the 2004 banking stability report of the European Central Bank ECB which may offer a deeper understanding on the 'numerator effect' of the liquidity risk measure ($C.LR_k$) applied in this research. The report finds that European banks seem to have relied more on money market funding than customer deposits as "the need to rely more on market funding has contributed negatively to banks' net interest income in the past few years. The funding gap[60] between loans granted to the non-bank sectors and deposits taken from these sectors has been positive in the last few years" (ECB, 2004a, p 10). Increasing funding gap suggests increasing reliance on market funding as opposed to customer deposits which in turn suggest higher liquidity risk (i.e. higher $C.LR_k$**)**. The ECB shows that customer funding gap was lowest in 2008 of around 10%, which then peaked in 2012 (nearly 22%) but dropped slightly in 2015 to just under 20% (ECB, 2004a, p 13).

Lastly, summary statistics for country-specific insolvency risk factor ($C.InsR_k$**),** calculated as the country-averaged ratio of equity capital to total assets, indicate that European banking systems held a core capital ratio at an average of nearly 9.5% over 2008 – 2018, which is slightly above the minimum capital requirements proposed by Basel I. This could reflect a somehow low risk of insolvency from a regulatory perspective. This in fact is confirmed by the ECB's 2010 bank stability report which found that "EU-15 banks' capital adequacy levels improved on average in 2012, as indicated by the increased overall solvency[61] and Tier 1 ratios…where overall

[60] Customer funding (deposits) is defined as non-bank deposits which include deposits from non-financial institutions, government and households (ECB, 2004a, p. 10). Market funding includes "issuance of debt securities such as medium-term notes, repos and unsecured interbank borrowing" (ECB, 2004a, p. 10).
Customer funding gap = (customer loans – customer deposits) / customer loans. Positive gap requires the bank to secure additional funding to support its lending operations as its sole core deposits base would not suffice.
[61] Overall solvency is represented in ECB's 2004 report by Total capital ratio which is (debt + equity capital) / total assets.

solvency ratio shifted towards the higher buckets, further indicating a strengthening in the solvency of EU-15 banks" (ECB, 2004a, p 15).

5.1.8 Variable Specification 2: Inefficiency Explanatory Variables

Since part of this chapter's contribution is to investigate the potential correlates (or z variables) of the technical inefficiency term u_{it} (Battese and Coelli, 1995), Table 3 below provides definitions and summary statistics for each of the proposed explanatory variables using data over 2008 – 2018 with 541 observations.

Table 3: Inefficiency Explanatory Variables

Bank-specific Variables	Description	Mean	Std.	Min.	Max.
CRA	**Credit Risk Appetite (ex ante)** = RWA[62] / Total Assets	.5672538	.1923314	.3000077	.9061569
$CrdtRsk$	**Credit Risk (ex post)** = Loan Loss Provisions / Total Loans	.0068087	.0076441	-.0027105	.0636142
$TrdgRsk$	**Trading Risk** Net Trading Revenue (trading income – trading costs)/ Total Revenues	.0382202	.0525289	-.0356843	.4288608
$InslvyRsk$	**Insolvency Risk** Equity Capital / Total Assets	.0829675	.0410361	.013538	.1559656
$LqdtyRsk$	**Liquidity Risk** = (Customer Deposits + SR Funding) / Liquid Assets	22.39321	13.73667	.0012587	95.72
$NetInt$	**Income Structure** Net Interest Revenue / Total Revenue	.3061708	.12194	.0466709	.5748032
$Funding$	**Funding Structure** = Customer Deposits / Total Funding	0.83053	0.179414	0.069534	.9470765
OBS	**Off-Balance Sheet ratio** = Credit-equivalent of OBS items/ Total Assets	.2868564	.3516229	.0001681	2.789446
$\frac{1}{2}\ln(Assets)^2$	**A non-linear form of Total Assets[63]**	62.8225	19.18659	23.75159	97.07446
t	**Linear form of time**	-	-	1 (2008)	10 (2018)
$\frac{1}{2}t^2$	**Non-linear form of time**	-	-	0.5	12.5

[62] RWA stands for Risk Weighted Assets.
[63] The point of regressing the inefficiency term against this non-linear expression of assets was to account for the size-effect on efficiency. It is worth noting that other non-linear forms of Assets were tested the form used in this research is the only one with which the model could converge. Lastly, specifying linear and non-linear forms of assets simultaneously in the function was not possible because the model could not converge even after 16000 iterations, for that reason the above –specified non-linear form of assets was applied.

The discussion below provides a brief reflection on some of the bank-specific variables explaining inefficiency as provided in Table 3.

To start with, summary statistics for credit risk appetite factor (CRA) suggest that European banks had about 57 % of these assets as risk weighted assets (RWA) in compliance with Basel I over the study period 2008 – 2018. CRA is introduced in the analysis as an ex ante measure of credit risk which provides –from a regulatory prospective– a pre-emptive measure for credit risk before the risk of default materializes. In terms of interpretation, the higher the CRA ratio the higher the level of RWA relative to the bank's total assets, which suggests that the bank is taking more credit risk from an ex ante perspective.

On the other hand, $CrdtRsk$ provides an ex post perspective on credit risk where provisions are built up on the basis of the perceived default risk on loans. The ratio of LLPs to total loans is typically investigated in efficiency literature as a proxy for credit risk. Statistics shows that loan loss provisions (LLPs) seem to account for 1.7% of total loans, which probably reflects a well-managed credit risk by European banks in general over 2010 – 2018. Nevertheless, the dispersion of $CrdtRsk$ reflects quite distinctive differences in exposure to credit risk by the sample banks, which makes it an interesting issue to investigate the impact of on technical efficiency. In terms of interpretation, the higher the $CrdtRsk$ ratio the greater the credit risk a given bank takes from an ex post perspective, which basically reflects the bank's assessment for the scale of default on the loan portfolio.

Bank-specific trading risk $TrdgRsk$, which is calculated as the ratio of net trading revenue to total revenues, reflects the risk-return payoff related to trading activities such that the larger the positive value of $TrdgRsk$ (i.e. the larger the value of net trading revenue proportionate to total revenues) the better the payoff. Statistics show that the average value for $TrdgRsk$ is about 4%, meaning that over 2008 – 2018, average European bank in this sample made about 4% of its total revenues from trading. The drop-in trading and investment banking activities due to the economic slowdown in the EU-15 and the associated problems in capital markets early this decade.

Bank-specific liquidity risk ($LqdtyRsk$) is calculated as the ratio of customer and short-term funding to liquid assets. Statistics show that $LqdtyRsk$ has an average of about 22, suggesting that the average value of customer and short-term funding is about 22 times the value of liquid assets held by European banks over 2010 – 2018. In other words, liquid assets accounted for about 4.6% of customer deposits and short-term funding. Higher value of $LqdtyRsk$ reflects greater exposure to liquidity risk and vice-a-versa.

For income structure as represented by ($NetInt$),calculated as the ratio of net interest income to total income, suggests that this sample of European banks seem to have generated, on average, around 30% of their total revenues from interest-related activities over the study period. It should be noted however that over the study period of 2010 -2018, the slowdown in the Euro area was reflected in reduced demand for credit which negatively impacted net interest income. Specifically, annual lending growth for the corporate sector dramatically declined from around 11% in 2010 down to nearly 4% in 2018. Equally, annual loan growth to households dropped from about 8% in 2010 to around 6% in 2018 (ECB, 2018). This translated in sluggish growth in retail interest margins. The changes that influenced net interest income constitute an interesting case to investigate the impact of on profit and cost efficiencies of European banks.

Concerning the funding structure ($Funding$), it seems that large European banks were heavily dependent on customer deposits given the average value of $Funding$ ratio of about 83%. Statistics show a considerable dispersion amongst this sample of European banks (with minimum and maximum of 7% and 94% respectively) in terms of their reliance on customer deposits to support their assets. Such considerable dispersion constitutes an interesting case to investigate in relation to both profit and cost efficiencies of the sample banks.

With regards to OBS variable (calculated as the ratio of the credit-equivalent of off-balance sheet items to total assets), statistics show an average of OBS ratio of around 30% (.2868564). This ratio suggests that around 1/3 of the assets on the balance sheet of an average European bank was held in OBS items over 2010 – 2018. This may have some implications on the profit and cost efficiencies as will be shown later.

The discussion now turns to exploring the preferred cost and profit functional forms to be estimated. The following sections will test the different modifications applied to both profit and cost functions from the statistical and regulatory conditions perspectives. The rationale behind modifying the functional form of the efficient frontier is twofold: first is to address the risk of estimating a functional form that does not fit the observed data (which parametric approaches are often criticized for in DEA literature), and second is to test for the impact of the risk factors. The modification process involves two main parts: the first is to modify the functional form itself to see whether the departure from the Translog form to more flexible forms is justified, and the second is to risk-modify the resulting functional form from the first step. The aim is to uncover the preferred models to be estimated and therefore to generate the corresponding profit and cost efficiencies for this sample of European banks. This in fact constitutes the main contribution of this empirical chapter.

5.2 Basic Translog Model 'T-Basic'

The starting point is to specify the basic form of Translog function which contains standard Translog terms (input prices, output quantities and their interactive terms) in addition to time dummies and their interactions with input prices and outputs. The result is the following Translog basic or (T-Basic) form:

$$
\begin{aligned}
&\ln(TC_i) or \ln(P_i) = \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n \\
&+ \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i + \rho t + \zeta \tfrac{1}{2} t^2 \\
&+ e_i
\end{aligned}
$$

(49)

Following Berger and Mester (1997 and 2003), the alternative profit and cost frontier functions have identical specification but with obviously different dependent variables and with different treatment for the composite error term e_i. The composite error term under the profit functional specification is the result of subtracting the random error component (ε_i) from the inefficiency component (u_i) such that e_i = ($\ln u_i - \ln \varepsilon_i$). Thus, the profit efficiency scores can take values up to unity which means that profit inefficient observations can only lie beneath the efficient frontier. On the other hand, under the cost function specification e_i = ($\ln u_i + \ln \varepsilon_i$) yielding cost efficiency scores with a minimum value of unity. This means that cost inefficient observations can only lie above the efficient frontier.

As can be seen, Translog function involves interaction between funding costs lw_1 and fixed costs lw_2 (which are normalized by the price of labour) with the different three outputs forming the terms $\sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn}\ln w_j \ln y_n$. The rationale for including the interactive terms of funding costs with outputs, for instance, is to reflect the variations of funding costs (lw_1) - as a percentage of labor expenses – over the business cycle of outputs. These variations can affect the mix of funding liabilities between retail deposits and money market or interbank funding (Humphrey and Vale, 2004). Such interaction is empirically captured by the coefficients of these interactive input prices-output quantities terms. The same logic applies to the interaction between fixed capital costs (lw_2) and output quantities.

Now, before embarking on modifying the basic form of the Translog function (T-Basic) any further, it is necessary to test the initial appropriateness of the Translog form given this research data. This is delivered by subjecting T-Basic to three main tests: Homotheticity, Cobb-Douglas, and Non-jointness as shown in the next section.

5.2.1 Testing the Appropriateness of 'T-Basic'

Certain structural tests should be applied in order to decide on the initial appropriateness of the Translog functional form following Kolari and Zardkoohi (1990).

According to Altunbas et al (1996), these tests include: Homotheticity, Cobb-Douglas, and Non-jointness. The notion of these structural tests is to impose certain restrictions on the Translog function defined in (49) page 202 and then compare the restricted against the unrestricted models using F tests. The following table summarizes the results of the three tests.

Table 4: Structural test results

Test Performed	**Chi2**	**Degrees of Freedom**	**P-value**	**Decision at the 1% level**
Homotheticity	86.11	6	0.0000	Reject
Cobb-Douglas	1.1e+08	15	0.0000	Reject
Non-jointness	689.74	6	0.0000	Reject

The Homotheticity restriction implies that the cost function is separable in outputs and input prices (Altunbas and Molyneux, 1996). In terms of the Translog function T-Basic specified above, such separability of the cost function requires that coefficients τ_{jn} are significantly not different from zero for all j, n. Results in Table 4 show that the homotheticity restriction is strongly rejected at 1% level suggesting that the outputs-input prices interactive terms are statistically significant and therefore should be included in the Translog function.

The Cobb-Douglas restriction implies testing whether all second-order parameters introduced by the Translog form are insignificantly different from zero. In terms of the T-Basic cost function defined above, this restriction entails that: ν_{jn} = 0, γ_{jn} = 0, and τ_{jn} = 0. The Cobb-Douglas test is firmly rejected at the 1% level, suggesting the statistical significance of all second-order Translog terms hence confirming the appropriateness of a more flexible cost function under the Translog specification compared to the restrictive specification of Cobb-Douglas.

Finally, the Non-jointness restriction tests the hypothesis that banks under study have a separate production function for each output, i.e. the banking production technology implies non-joint production process. In terms of Translog form specified, non-jointness implies that v_{jn} are significantly not different from zero for all j, n. The non-jointness restriction is strongly rejected at the 1% level, indicating that the production process cannot be separated for each output. In other words, the marginal cost of a given output depends on the level of production of the other output, suggesting that cost savings can only be realized if outputs are jointly produced.

All the above-mentioned structural tests for the cost function specified by Translog form in (46) are decisively rejected at the 1% critical level. Therefore, the conclusion is that, given this research data set, there is no support for homothetic, Cobb-Douglas, or non-joint cost function. This result accordingly suggests that the more appropriate functional form to represent this sample banks' cost structure and production technology is a multi-input and multi-output Translog functional form.

With these results in mind, it is now safe to discuss the modification process of the Translog functional forms for profit and cost-efficient frontiers. The model T-Basic defined in (49) page 202 is therefore split into two functional forms: profit and cost-efficient frontier models. This step is necessary to take now because the two models have different dependent variables and the tests for regulatory conditions which will be applied on each form later necessitate this separation. The resulting functional forms are denoted as **T-Basic-C** when the dependent variable total costs ($\ln TC_i$) is specified in equation (50) and **T-Basic-P** when the dependent variable profits ($\ln P_i$) is specified in equation (51). This separation is necessary because the corresponding log-likelihood functions will have different values accordingly.

In the next section, these two basic Translog forms will be tested against more advanced Translog forms **T-C** (Translog cost function) and **T-P** (Translog profit function) to decide whether it is justified to depart from the T-Basic functional forms.

5.2.2 Translog Cost function 'T-C'

Variables specified under **T-C** function include those of **T-Basic** plus EU-membership and Euro-adoption dummies denoted as EU and $EURO$ respectively, in addition to a set of country-specific macroeconomic variables including: GDP growth (GDP), inflation rate (π_k), short-term interest rates (r_k), bank-intermediation measure ($M2_k$), and a loan-market concentration index ($HERF_k$). These additional variables are all defined in Table 2. The introduction of these variables is vital to capture the potential impact of country-specific differences on cost efficiency. The resulting Translog cost functional form T-C is therefore specified as:

$$
\begin{aligned}
&\ln(TC_i) = \alpha + \sum_{i=1}^{2} \phi_i \ln w_i + \sum_{i=1}^{3} \varphi_i \ln y_i \\
&+ \tfrac{1}{2} \sum_{j=1}^{2} \sum_{n=1}^{2} \gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2} \sum_{j=1}^{3} \sum_{n=1}^{3} \nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2} \sum_{n=1}^{3} \tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2} \psi_i t \ln w_i + \sum_{i=1}^{3} \omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + [\beta_1 Eu_i + \beta_2 Euro_i + \beta_3 GDP_i + \beta_4 \pi_i \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k] \\
&+ (\ln \varepsilon_i + \ln u_i)
\end{aligned}
$$

(50)

To assess the statistical significance of the added variables under the T-C specification, two tests are applied: the log-likelihood ratio and Wald Chi-squared tests. Both tests achieve the same purpose of assessing the joint significance of the added variables however both are used simply to provide alternative ways of verifying the validity of the functional form modification process. To conduct these tests, the T-Basic-C and T-C models are estimated using Maximum Likelihood Estimation (MLE) technique so as to evaluate the corresponding log-likelihood functions for the two models and evaluate if improvements to the values of these functions are statistically significant. The results of these two tests are presented in Table 5 below.

Table 5

Cost Function	Number of Variables	Log-likelihood Value	LR Chi2 (7)[64]	Wald Chi2 (7)	Description
T-Basic-C	27	269.8357	31.02**	42.23**	T-Basic-C nested in T-C
T-C	34	291.847	-	-	-

It is clear that the improvement in the value of the log-likelihood function from 269.8357 to 291.847 is statistically significant at 1% level. This is brought about by the 7 added dummies and macroeconomic variables in the T-C model. Also, both test statistics show that the H0 stating the joint insignificance of the (7) added variables can comfortably be rejected at 1% critical level. Therefore, the departure from the basic Translog cost function T-Basic-C to the Translog cost function T-C is well justified.

5.2.3 Translog Profit function 'T-P'

Similar methodology to that of the previous section is applied to test whether the departure from the basic Translog profit function T-Basic-P to the more flexible Translog profit function T-P is justified. The seven added variables in the Translog profit function T-P specified in (51) were defined earlier in the previous section, with the difference between T-P and T-C being that the dependent variable under T-P is profits instead of total costs. Accordingly, the Translog profit functional form is specified as follows:

[64] (**) and (*) indicate that P-value > Chi2 at 1% and 5% significance levels respectively. This notation applies throughout the chapter. The value in parenthesis refers to the number of additional variables in the unrestricted model being tested.

$$
\begin{aligned}
\ln(P_i) &= \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 Eu_i + \beta_2 Euro_i + \beta_3 GDP_i + \beta_4 \pi_i \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ (\ln u_i - \ln \varepsilon_i)
\end{aligned}
$$

(51)

Table 6 below investigates whether adding these seven variables to improve the T-Basic-P is justified.

Table 6

Profit Function	Number of Variables	Log-likelihood Value	LR Chi2 (7)	Wald Chi2 (7)	Description
T-Basic-P	27	173.7211	21.25**	22.50**	T-Basic-P nested in T-P
T-P	34	197.3437	-	-	-

Table 6 shows that the improved value of the estimated log-likelihood function under the T-P specification from 173.7211 to 197.3437 is statistically significant at 1% critical level. Both test statistics show that the H0 stating the joint insignificance of the (7) added variables can comfortably be rejected at 1% critical level. Thus, the departure from **T-Basic-P** to **T-P** is well supported.

5.2.4 Fourier Flexible Cost Function 'F-C'

In this section a set of Fourier Flexible terms is introduced in an attempt to overcome the restrictions of the Translog specification (McAllister and McManus, 1993). Specifically, first-order Fourier terms containing both rescaled input prices and output quantities are added. The Fourier series is restricted to first order terms in order to avoid the degrees of freedom problem given this research sample size.

The introduction of the Fourier terms as such has proven to be essential to meet the cost function's regulatory conditions in a satisfactory manner. The specification of Fourier Flexible Cost function **F-C** is similar to that of **T-C** as the latter is nested in the former given the (10) added first order Fourier terms as shown in equation (48) below[65]:

$$
\begin{aligned}
\ln(TC_i) &= \alpha + \sum_{i=1}^{2} \phi_i \ln w_i + \sum_{i=1}^{3} \varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2} \gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3} \nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3} \tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2} \psi_i t \ln w_i + \sum_{i=1}^{3} \omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_{ki} + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5} [\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ (\ln \varepsilon_i + \ln u_i)
\end{aligned}
$$

(52)

Table 7 below investigates whether adding these ten variables to improve the T-C is justified.

Table 7

Cost Function	Number of Variables	Log-likelihood Value	LR Chi2 (10)	Wald Chi2 (10)	Description
T-C	34	291.847	61.07**	72.22**	T-C nested in F-C
F-C	44	314.3805	-	-	-

[65] Altunbas et al (2000) provides an excellent summary of the arguments for using the Fourier as opposed to the Translog specification. They point out that "the Fourier Flexible functional form should be preferred over the Translog because the former better approximates the underlying cost function across a broad range of outputs. The semi-non-parametric Fourier functional form has desirable mathematical and statistical properties because an (infinite) Fourier series is capable of representing any function exactly, and even truncated Fourier series can approximate a function reasonably well throughout its entire range. In contrast, when using parametric methods like the Translog, one holds the maintained hypothesis that the bank industry's true cost function has the Translog form. If this maintained hypothesis is false misspecification error occurs. When using the Fourier functional form, one avoids holding any maintained hypothesis by allowing the data to reveal the true cost function through a large value of fitted parameters." (p 1609)

Table 7 shows that the enhanced value of the estimated log-likelihood function under the F-C specification from 291.847 to 314.3805 is statistically significant at 1% level. Both test statistics show that the H0 stating the joint insignificance of the (10) added Fourier terms can comfortably be rejected at 1% critical level. Thus, the departure from **T-C** functional form to **F-C** is well supported.

5.2.5 Fourier Flexible Profit function 'F-P'

The specification of Fourier Flexible Profit frontier **F-P** is an augmented form of Translog Profit frontier **T-P**. The latter is nested in the former because of the added first order Fourier terms. The functional expression of F-P is therefore defined as follows:

$$
\begin{aligned}
\ln(P_i) &= \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_{ki} + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5}[\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ (\ln u_i - \ln \varepsilon_i)
\end{aligned}
\tag{53}
$$

Table 8 below investigates whether adding these ten variables to improve the T-P is justified.

Table 8

Profit Function	Number of Variables	Log-likelihood Value	LR Chi2 (10)	Wald Chi2 (10)	Description
T-P	34	197.3437	40.43**	45.14**	T-P nested in F-P
F-P	44	204.5602	-	-	-

Table 8 shows that the increase in the value of the estimated log-likelihood function under the F-P specification from 197.3437 to 204.5602 is statistically significant at 1% level. Both test statistics show that the H0 stating the joint insignificance of the (10) added Fourier terms can comfortably be rejected at 1% critical level. Thus, the departure from T-P to F-P is well supported.

The second stage of the modification process involves incorporating risk factors into the frontier functional form. A set of four country-specific risk proxies for trading, credit, liquidity, and insolvency risks are introduced which are denoted as $C.TR$, $C.CR$, $C.LR$, and $C.InsR$ respectively. These risk factors are basically bank-specific that are averaged for each group of banks operating in a specific country yielding the country-specific risk factors. All of the risk factors are defined in Table 1.

Prior to introducing these risk factors, it is necessary to test for potential multicollinearity amongst these variables by establishing the correlation coefficient matrix showed in Table 9 below.

Table 9

	$C.TR$	$C.CR$	$C.LR$	$C.InsR$
$C.TR$	1			
$C.CR$	0.0176	1		
$C.LR$	-0.1109	0.1936	1	
$C.InsR$	0.009	0.2967	0.366	1

Results in Table 9 show little evidence of multicollinearity amongst the four risk variables. Nonetheless, the correlation between country-specific liquidity risk ($C.LR$) and insolvency risk ($C.InsR$) might be suspected as problematic. Such doubt is clarified by estimating both profit and cost-efficient frontiers twice, i.e., with and without one of the suspected variables, so as to examine whether this would alter the statistical significance of the remaining variable. To this end, country-specific liquidity risk $C.LR$ is chosen as the dropped variable. The cost and profit frontiers, F-C and F-P, are accordingly specified and each is twice estimated with and without $C.LR$. The results of this testing procedure are shown in Table 10 below.

Table 10: Incorporating Country-specific Insolvency Risk

Frontier Specification	Variable Specification	Results for Country-specific Insolvency Risk ($C.InsR$)			
		Coeff.	Std. Err.	z	P|z|
Cost Frontier	With $C.LR$	.2738111	.1027331	2.67	0.008
	Without $C.LR$	.3283691	.1253198	2.62	0.009
Profit Frontier	With $C.LR$	-.1461599	.0808376	1.81	0.071
	Without $C.LR$	-.1000841	.0715507	1.40	0.162

Under the cost frontier specification, it is clear that the suspected high correlation between country-specific liquidity and insolvency risks, $C.LR$ and $C.InsR$, is having no significant effect for the estimation results of $C.InsR$ when the model is estimated with and without the other risk variable, $C.LR$. The impact of insolvency risk $C.InsR$ remains significant and positive at 1% critical level although changing slightly in magnitude. The standard error of $C.InsR$ is also not showing dramatic increase. Likewise, under the profit frontier specification, estimation results of $C.InsR$ are also not showing dramatic change. $C.InsR$ remains negative and statistically insignificant at 1% levels when the profit model is estimated with and without introducing $C.LR$. The standard error of $C.InsR$ is also not showing any dramatic change.

In light of the results above, it is therefore statistically safe to estimate profit and cost-efficient frontiers with the two risk variables simultaneously accounted for. This paves the way for further modification to be applied to the Fourier Flexible cost **F-C** and profit **F-P** models discussed above in terms of introducing the proposed four country-specific risk variables: trading, credit, liquidity, and insolvency risks. This will accordingly yield two new models: The Risk-modified Fourier Flexible Cost model **R-F-C** and the Risk-modified Fourier Flexible Profit model **R-F-P**.

5.2.6 Risk-Modified Fourier Flexible Cost Model 'R-F-C'

The Risk-modified Fourier Flexible Cost model R-F-C is an augmented form of the Fourier Flexible Cost model F-C function defined in ((54)). This is achieved by incorporating the four country-specific risk variables including: trading, credit, liquidity, and insolvency risks. The resulting R-F-C is defined as:

$$
\begin{aligned}
\ln(TC_i) &= \alpha + \sum_{i=1}^{2} \phi_i \ln w_i + \sum_{i=1}^{3} \varphi_i \ln y_i \\
&+ \tfrac{1}{2} \sum_{j=1}^{2}\sum_{n=1}^{2} \gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2} \sum_{j=1}^{3}\sum_{n=1}^{3} \nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3} \tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2} \psi_i t \ln w_i + \sum_{i=1}^{3} \omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_{ki} + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5} [\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ \varpi_i C.TR_k + \upsilon_i C.CR_k + \tau_i C.LR_k + \varsigma_i C.InsR_k \\
&+ (\ln \varepsilon_i + \ln u_i)
\end{aligned}
\tag{54}
$$

To test the validity of this modification, log-likelihood and Wald Chi-squared tests are applied as shown in Table 11below:

Table 11

Cost Function	Number of Variables	Log-likelihood Value	LR Chi2 (4)	Wald Chi2 (4)	Description
F-C	44	314.3805	43.27**	48.68**	F-C nested in R-F-C
R-F-C	48	336.0138	-	-	-

(**) refer to statistical significance at 1% level respectively.

Clearly, the improvement in the value of the log-likelihood function from 314.3805 to 336.0138 is statistically significant at 1% level. This is caused by the added (4) country-specific risk factors in R-F-C. Both test statistics show that the H0 stating the joint insignificance of the (4) added risk factors can comfortably be rejected at 1% critical level. The conclusion therefore is that the four risk factors are jointly significant in improving the model's fit to the data. Accordingly, the departure from the Fourier Flexible Cost function **F-C** to the Risk-modified Fourier Flexible Cost function **R-F-C** is well justified.

As a final check, the two structural tests are conducted on the most restrictive model of Basic Translog Cost function (T-Basic-C) and the preferred Risk-modified Fourier Flexible Cost function (R-F-C). This is to further ensure the validity of the preferred R-F-C model before conducting the estimation. The results of these tests are presented in Table 12 below.

Table 12

Cost Function	Number of Variables	Log-likelihood Value	LR Chi2 (21)	Wald Chi2 (21)	Description
T-Basic-C	27	269.8357	132.36**	192.52**	T-Basic-C nested in R-F-C
R-F-C	48	336.0138	-	-	-

(**) refer to statistical significance at 1% level respectively.

The dramatic increase in the value of the log-likelihood function brought about by the additional 21 variables is statistically significant. This is confirmed by the statistical significance of both test statistics in rejecting the null hypothesis at the 1% level. Hence the 21 added variables are jointly significant. Therefore, the conclusion that can be derived is this: the R-F-C model is strongly preferred to the less flexible cost models tested as it offers a significantly enhanced representation (fit) to the data. Therefore, the **R-F-C** model is estimated to analyse European banks cost efficiency.

5.2.7 Risk-Modified Fourier Flexible Profit Model 'R-F-P'

The Risk-modified Fourier Flexible Profit model R-F-P is a modified functional form of the Fourier Flexible Profit model F-P defined in (55). This is achieved by incorporating four country-specific risk variables including: trading, credit, liquidity, and insolvency risks. The resulting **R-F-P** is defined below:

$$
\begin{aligned}
\ln(P_i) &= \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_{ki} + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5}[\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ \varpi_i C.TR_k + \upsilon_i C.CR_k + \tau_i C.LR_k + \varsigma_i C.InR_k \\
&+ (\ln u_i - \ln \varepsilon_i)
\end{aligned}
$$

(55)

Table 13

Profit Function	Number of Variables	Log-likelihood Value	LR Chi2 (4)	Wald Chi2 (4)	Description
F-P	44	204.5602	14.49**	15.06**	F-P. nested in R-F-P
R-F-P	48	211.8043	-	-	-

The enhanced value of the log-likelihood function from 204.5602 to 211.8043 is statistically significant at the 1% level. This is caused by the added (4) risk factors in R-F-P. Both test statistics show that the H0 stating the joint insignificance of the (4) added risk factors can comfortably be rejected at 1% critical level. The conclusion therefore is that the four risk factors are jointly significant in improving the model's fit to the data. This suggests that the departure from the **F-P** to the **R-F-P** model is well justified.

As a final check, the two structural tests are conducted on the most restrictive model of Basic Translog Profit function (T-Basic-P) and the preferred Risk-modified Fourier Flexible Profit function (R-F-P). This is to further ensure the validity of the preferred R-

F-P model before conducting the estimation. The results of these tests are presented in Table 14 below.

Table 14

Profit Function	Number of Variables	Log-likelihood Value	LR Chi2 (21)	Wald Chi2 (21)	Description
T-Basic-P	27	173.7211	132.36**	192.52**	T-Basic-P nested in R-F-P
R-F-P	48	211.8043	-	-	-

This is confirmed by the statistical significance of both test statistics in rejecting the null hypothesis at the 1% level hence the 21 added variables are jointly significant. Therefore, the conclusion that can be derived is this: R-F-P model is strongly preferred to the less flexible profit models tested as it offers a significantly enhanced representation (fit) to the data, therefore the **R-F-P** model is estimated to analyse European banks profit efficiency.

Having decided on the preferred profit and cost frontier models, the preferred cost model is tested for regulatory conditions. Despite the similar specification of the profit and cost functions, the literature implies that the profit function does not have to be tested for regulatory conditions as does the cost function (Diewert and Wales, 1987 and Berger and Mester, 1997). Testing the regulatory of the preferred cost function (R-F-C) is the subject of the next section.

5.3 Estimating the Preferred Profit and Cost Models

Having explored the different specifications of the preferred model and tested the regularity conditions of each, this led to deciding on estimating the Risk-modified Cost and Profit frontier functions, namely, R-F-C and R-F-P. These models are defined again in equations below. These two models are used to produce the corresponding efficiency estimates. The Risk-modified Cost Frontier **R-F-C** is specified as:

$$
\begin{aligned}
\ln(TC_{it}) &= \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_k + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5}[\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ \varpi_i C.TR_k + \upsilon_i C.CR_k + \tau_i C.LR_k + \varsigma_i C.InR_k \\
&+ (\ln \varepsilon_i + \ln u_i)
\end{aligned}
$$

(56)

And the Risk-modified Profit Frontier **R-F-P** is specified as:

$$
\begin{aligned}
\ln(P_{it}) &= \alpha + \sum_{i=1}^{2}\phi_i \ln w_i + \sum_{i=1}^{3}\varphi_i \ln y_i \\
&+ \tfrac{1}{2}\sum_{j=1}^{2}\sum_{n=1}^{2}\gamma_{jn} \ln w_j \ln w_n + \tfrac{1}{2}\sum_{j=1}^{3}\sum_{n=1}^{3}\nu_{jn} \ln y_j \ln y_n \\
&+ \sum_{j=1}^{2}\sum_{n=1}^{3}\tau_{jn} \ln w_j \ln y_n + \sum_{i=1}^{2}\psi_i t \ln w_i + \sum_{i=1}^{3}\omega_i t \ln y_i \\
&+ \rho t + \zeta \tfrac{1}{2} t^2 + \beta_1 EU_k + \beta_2 EURO_k + \beta_3 GDP_{ki} + \beta_4 \pi_k \\
&+ \beta_5 r_k + \beta_6 M2_k + \beta_7 .HERF_k \\
&+ \sum_{n=1}^{5}[\phi_n \cos(x_n) + \omega_n \sin(x_n)] \\
&+ \varpi_i C.TR_k + \upsilon_i C.CR_k + \tau_i C.LR_k + \varsigma_i C.InR_k \\
&+ (\ln u_i - \ln \varepsilon_i)
\end{aligned}
$$

(57)

Where k Specifies the number of countries in the panel k = 1, 2… 14. The two frontiers are estimated using the Stochastic Frontier Analysis (SFA) following the single-stage estimation approach where the explanatory variables of the inefficiency term are simultaneously estimated with the frontier functional form. The literature postulates two mainly applied models for specifying the technical inefficiency model: Battese and Coelli (1992) and Battese and Coelli (1995). The inefficiency term's model adopted in this research is the time-flexible model of Battese and Coelli (1995). The two models were introduced in the methodology chapter.

Estimating the more time-flexible model of Battese and Coelli (1995) is preferred to the (1992) model for two reasons. First, the (1995) model permits the application of single-stage estimation procedure. Second, the (1995) model does not impose a time trend on inefficiencies. For these reasons, Battese and Coelli (1995) inefficiency model is adopted.

Because the values of the inefficiency term effects u_{it} are unobservable, but the values of ε_{it} are, u_{it} is therefore predicted conditional on ε_{it} estimates. This is achieved by applying the conditional expatiation of u_{it} on the observed value of ε_{it} (Coelli, 1996). Accordingly, observation-specific predicted inefficiencies are obtained as $E(u_{it} \mid \varepsilon_{it})$.

Once predicted inefficiencies u_{it} are obtained, bank-specific inefficiency scores (EFF_{it}) are produced as (Coelli, 1996, p 8):

$$EFF_{it} = E(TC_{it}^{*} \mid u_{it}, x_{it}) / E(TC_{it}^{*} \mid u_{it} = 0, x_{it})$$

(58)

Equation (54) produces cost efficiency scores, where TC_{it}^{*} is the actual total cost of the bank i at time period t, which effectively will be equal to $\exp(TC_{it})$ and u_{it} will be expressed as $\exp(u_{it})$ as the dependent variable in the frontier function is logged. x_{it} are the frontier variables of the R-F-C cost model. Cost efficiency scores take the values between one and infinity. On the other hand, producing profit efficiency scores follows similar methodology to that of equation ((56)) with profits P_{it}^{*} being specified instead of total costs TC_{it}^{*}. Profit efficiency scores take the values up to unity.

Predicted inefficiencies u_{it} can then be regressed against a set of explanatory (z_{it}) variables in order to explore the drivers of technical inefficiency, and to allow for time-variancy, both linear and non-linear forms of the time dummy are specified as z_{it} variables as shown in equation (59) below. Following Battese and Coelli (1995, p 328), the inefficiency term u_{it} is modelled against a set of bank-specific variables such that:

$$u_{it} = \alpha + \beta_1 CRA_{it} + \beta_2 CrdtRsk_{it} + \beta_3 TrdgRsk_{it} + \beta_4 InslvcyRsk_{it} + \beta_5 LqdtyR_{it} + \beta_6 NetInt_{it} + \beta_7 Funding_{it} + \beta_8 OBS_{it} + \beta_9 \tfrac{1}{2}\ln(Assets_{it})^2 + \beta_{10} t_t + \beta_{11} \tfrac{1}{2} t_t^2$$

(59)

Where i is the number of banks in the panel: i=1, 2, 3… 541, and t represents time periods hence takes the values of 1, 2… 5. A full description of these bank-specific variables can be found in the table below.

Before estimating the profit and cost frontier models specified above in a single-stage process, it is important to check the pair-wise correlations amongst the z_{it} variables to detect any possible multicollinearity. To this purpose, the correlation coefficient (variance-covariance) matrix is produced as presented below:

Table 15: Variance-Covariance matrix for bank-specific Risk factors

	CRA	$CrdtRsk$	$TrdgRsk$	$LqdtyRsk$	$InslvcyRsk$	OBS	$NetInt$	$Funding$	$\frac{1}{2}\ln(Assets)^2$
CRA	1								
$CrdtRsk$	0.10 (0.1564)*	1							
$TrdgRsk$	-0.09 (0.5418)	-0.07 (0.2897)	1						
$LqdtyRsk$	0.03 (0.6389)	-0.05 (0.4909)	0.13 (0.0589)	1					
$InslvcyRsk$	0.31 (0.0115)	-0.06 (0.3808)	0.23 (0.2341)	0.18 (0.0115)	1				
OBS	0.13 (0.0680)	0.16 (0.0269)	-0.10 (0.1346)	-0.05 (0.4597)	-0.06 (0.3741)	1			
$NetInt$	0.24 (0.0231)	0.27 (0.1848)	-0.22 (0.3301)	-0.00 (0.9601)	0.32 (0.0199)	0.31 (0.7341)	1		
$Funding$	-0.01 (0.7942)	0.09 (0.0298)	0.31 (0.1004)	0.07 (0.3156)	0.18 (0.0309)	0.28 (0.1082)	0.01 (0.8281)	1	
$\frac{1}{2}\ln(Assets)^2$	-0.30 (0.0301)	-0.27 (0.1231)	-0.07 (0.0324)	-0.12 (0.0943)	-0.22 (0.0123)	-0.08 (0.2598)	-0.25 (0.0427)	-0.07 (0.2341)	1

**Values in parentheses represent the p-values of the corresponding correlation coefficients*

Generally, all variables seem to be well-behaving in terms of collinearity as none of the correlation coefficients is greater than 0.5 (Newbold, 2007), and none of these correlations is statistically significant at the 1% level (although some are at the 5% level, however given the small size of the sample the 1% level is considered as the main criterion for correlation significance). Accordingly, it is statistically sound to include all of the above z_{it} variables in the inefficiency model. The following section discusses estimation results of the preferred cost model R-F-C and the determinants of the cost inefficiency term.

5.4 Empirical Results 1: Cost Function (R-F-C)

The discussion now turns onto analysing the empirical findings related to the estimated cost and profit frontiers and their corresponding inefficiencies. To this end, a common profit and cost-efficient frontiers are specified because of similarities in regulation and technology across European banking markets as illustrated by Berger et al (2001), Casu and Molyneux (2003), Casu et al (2004), and Barros et al (2007). McAllister and McManus (1993) reinforces the argument of fitting a single (common) frontier over the entire sample period on the basis of avoiding the degrees of freedom problem for parametric methods. Estimating a single frontier assumes that all countries (banking systems) in the sample are situated on the same frontier despite their differences. The potential downside of this assumption is mitigated by accounting for country-specific differences in terms of risk profiles and macroeconomic conditions as control variables in the frontier functional form.

5.4.1 Hypothesis Testing

Prior to conducting the estimation of the preferred cost model R-F-C and producing the corresponding cost efficiency estimates, there are three main tests that need to be applied to the model, these, according to Battese and Coelli (1995, p 330), are: technical inefficiency test, stochasticity test, and inefficiency correlates test. The aim of these tests is to ensure whether the cost model (1) has an inefficiency effect, (2) it is stochastic, and (3) if the proposed inefficiency term's correlates are jointly significant.

5.4.2 Technical Cost Inefficiency Test

This test investigates whether the cost model R-F-C has inefficiency effects or not. Coelli (1998) explains that this test examines if the inefficiency variance is significantly different from zero. This entails testing $H0: \sigma_u^2 = 0$ against the alternative $H1: \sigma_u^2 > 0$ since inefficiencies can only be nonnegative. The results of this test are resented in the table below:

Table 16: Composite error term components

Components of the composite error term	Coeff.	Std. Err.	z	P|z|
Variance of the Random Error component ($\ln \sigma_v^2$)	-24.32985	5.773881	-4.21	0.000
Variance of the Technical inefficiency component ($\ln \sigma_u^2$)	-5.472311	.1034219	-52.91	0.000

Findings clearly show that the null of absent inefficiency effects is strongly rejected since $\ln \sigma_u^2$ is found to be statistically very significant at the 1% level. Hence the R-F-C cost model has technical inefficiency effects.

5.4.3 Stochasticity Test

Having ensured the existence of the inefficiency effects, these effects will need to be checked for stochasticity. This is because applying the single-stage estimation approach requires that the inefficiency effects to be stochastic as assumed by the Battese and Coelli (1995) model. This will also have a direct impact on the choice of the estimation technique (OLS or SFA). Stochasticity test involves testing the significance of the parameter $\gamma = \sigma_u^2 / \sigma^2$ which implies testing whether the variance of the inefficiency term constitute a statistically significant proportion of the variation of the composite (or aggregate) error term.

Accordingly, the null states that the technical inefficiency is not stochastic, that is $H_0 : \gamma = 0$, which effectively suggests testing whether σ_u^2 is statistically different from zero in relation to σ^2. This is tested against $H_1 : \gamma > 0$ (as technical inefficiency is always assumed as being nonnegative) which implies that the frontier is stochastic. Not rejecting the null suggests that the frontier's parameters can consistently be estimated using ordinary least squares OLS where the inefficiency term can be removed from the model (Coelli, 1996, p 5).

To empirically perform this, according to Gutierrez et al (2001), the R-F-C cost model is estimated using MLE and subsequently using the log-likelihood ratio test (LR)[66]. Results show that the null hypothesis that σ_u^2= 0 is comfortably rejected at 1% critical level: LR statistics ($\bar{\chi}_{01}^2$) = 79.83 with P-value = 0.000. Therefore, the technical inefficiency model associated with the cost model R-F-C is stochastic because the inefficiency term has a significant variation in relation to the variation of the composite error term hence the model should be estimated using SFA.

5.4.4 Inefficiency Correlates Test

This test tests the null hypothesis that the inefficiency term's correlates are jointly insignificantly different from zero, although the individual effects of one or more of the correlates (determinants) may be statistically insignificant. Given the inefficiency model specified in equation (19(59)), $H_0 : \beta_1 = \beta_2 = ... = \beta_{11} = 0$. The test statistics is Chi-squared distributed with its parameter being equal to the parameters assumed to be zero. Results show that H_0 is comfortably rejected at the 1% critical level where the tests statistics chi2 (11) = 38.81 and Prob. > chi2 = 0.0002.

[66] The LR test here performs the test on the boundary of the parameter space since the tested parameter is always positive and the test examines whether the parameter is different from zero (Coelli, 1996). Gutierrez et al (2001, p 15) states that "In such cases, the limiting distribution of the maximum-likelihood estimate of the parameter in question is a normal distribution that is halved, or chopped-off at the boundary zero in this case. As a result the distribution of the LR test statistic is not the usual chi-square with 1 degree of freedom, but instead a 50:50 mixture of a chi-square with no degrees of freedom (i.e. a point mass at zero or normal distribution) and a chi-square with 1 degree of freedom". Having said that, the p-value will be equal 1 if LR statistic is close to zero, and one-half of the probability that Chi2 with 1 degree of freedom if p-value is greater than LR statistic.

All in all, the results of the above three tests indicate that the preferred cost model R-F-C has a technical inefficiency effects that is stochastic and that its suggested correlates are jointly significant. Having found that, the next step now is to estimate the R-F-C model as defined in equation (54) simultaneously with the inefficiency model defined in equation (19), according to the single-stage SFA approach.

5.4.4.1 Estimating the Cost Function (R-F-C)

To produce observation-specific technical cost inefficiency estimates, the cost frontier model, R-F-C that is defined in ((56)), and the technical cost inefficiency model defined in ((59)) are estimated under the single-stage SFA. Crucial to producing efficient parameters' estimates (lowest standard errors possible) is to ensure the function's convergence. Convergence is important to achieve since this is indicative of reaching the greatest value of the log-likelihood function (LL) which corresponds to the R-F-C model in the estimation process. Without achieving the convergence of (LL), the resulting parameters' estimates would be less reliable (or less efficient). In light of this, Figure 8 graphically displays the convergence of the (LL) function representing the estimated cost frontier which specifies the R-F-C model jointly with the inefficiency model.

Figure 8: Convergence of estimated R-F-C cost frontier with Battese & Coelli (1995), inefficiency model and half-normal distributional assumption specified

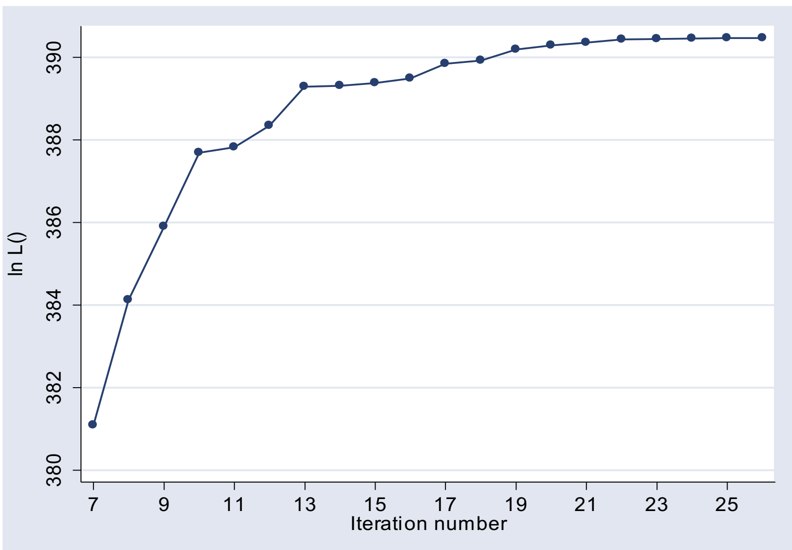


Figure 8 depicts the progression of the iterative process applied by the maximum likelihood estimation technique MLE to evaluate the LL function which corresponds to the R-F-C model. Convergence is achieved when the curve becomes eventually flat. Figure 4 shows that the cost function converges to its maximum possible value of 390.46699 after 26 iterations. The estimation outcome for the cost frontier model is shown below:

Iteration 24: log likelihood = 490.45501
Iteration 25: log likelihood = 490.46623
Iteration 26: log likelihood = 490.46699

Stoc. frontier normal/half-normal model

Number of obs	=	541
Wald chi2(48)	=	1.089e+09
Prob > chi2	=	0.0000

Log likelihood = **490.46699**

$\ln TC_{it}$	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
lw_1	.8540357	.1593414	5.36	0.000	.5417322	1.166339
lw_2	.3764798	.0950634	3.96	0.000	.1901588	.5628007
ly_1	-2.298756	.7925618	-2.90	0.004	-3.852148	-.7453634
ly_2	-.5442946	.7110157	-0.77	0.444	-1.93786	.8492705
ly_3	.2851808	.1043743	2.73	0.006	.080611	.4897507
$\frac{1}{2} lw_1 lw_1$	-.1531868	.0912211	-1.68	0.093	-.3319769	.0256033
$\frac{1}{2} lw_2 lw_2$	-.0308776	.0172508	-1.79	0.073	-.0646886	.0029333
$lw_1 lw_2$	.0022103	.0067954	0.33	0.745	-.0111085	.0155291
$\frac{1}{2} ly_1 ly_1$	1.416898	.3488817	4.06	0.000	.7331028	2.100694
$\frac{1}{2} ly_2 ly_2$	.9341489	.3470001	2.69	0.007	.2540413	1.614257
$\frac{1}{2} ly_3 ly_3$	-.123604	.0548469	-2.25	0.024	-.231102	-.016106
$ly_1 ly_2$	-.0368887	.0454413	-0.81	0.417	-.1259521	.0521747
$ly_1 ly_3$	.0236013	.0321541	0.73	0.463	-.0394196	.0866223
$ly_2 ly_3$	-.1172417	.0198196	-5.92	0.000	-.1560875	-.078396
$lw_1 ly_1$	.0122847	.0153713	0.80	0.424	-.0178425	.0424119
$lw_1 ly_2$	-.0667754	.020713	-3.22	0.001	-.1073722	-.0261787
$lw_1 ly_3$	.0411668	.0097096	4.24	0.000	.0221364	.0601972
$lw_2 ly_1$	-.0567643	.007637	-7.43	0.000	-.0717324	-.0417961
$lw_2 ly_2$	-.0512964	.0092335	-5.56	0.000	-.0693937	-.0331991
$lw_2 ly_3$	.0070236	.0047713	1.47	0.141	-.002328	.0163753
tlw_1	-.0020947	.0042805	-0.49	0.625	-.0104843	.0062949
tlw_2	.0086007	.0019827	4.34	0.000	.0047148	.0124867
tly_1	-.0099285	.006155	-1.61	0.107	-.0219921	.0021351
tly_2	-.0202139	.0048177	-4.20	0.000	-.0296565	-.0107714
tly_3	.0106774	.0053617	1.99	0.046	.0001687	.0211862
t	.0089875	.0403856	0.22	0.824	-.0701669	.0881419
$\frac{1}{2} t^2$	-.0055524	.0086188	-0.64	0.519	-.0224449	.0113402
$\cos x_1$	-.1752925	.0691752	-2.53	0.011	-.3108734	-.0397115

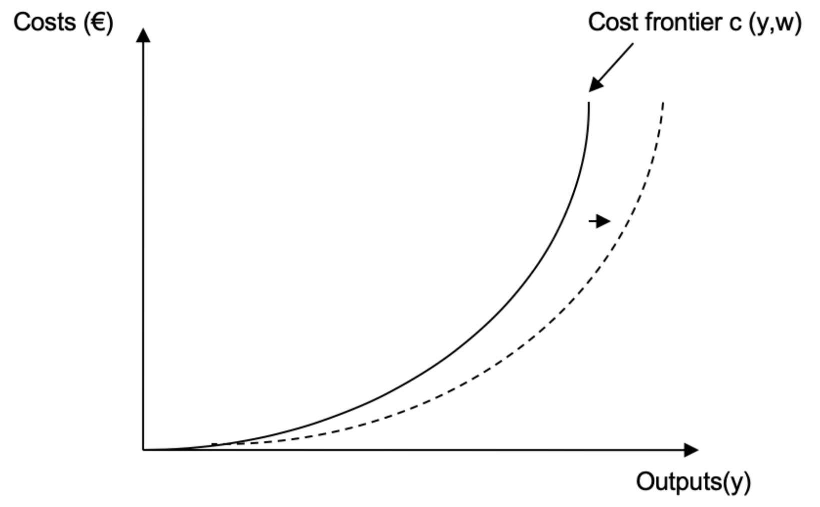
$\sin x_1$	-.3728306	.096756	-3.85	0.000	-.5624689	-.1831923
$\cos x_2$	-.2737374	.1975796	-1.39	0.166	-.6609863	.1135116
$\sin x_2$	.1599284	.1000625	1.60	0.110	-.0361906	.3560473
$\cos x_3$	.1392772	.0480462	2.90	0.004	.0451083	.233446
$\sin x_3$	.0415813	.0194783	2.13	0.033	.0034046	.0797579
$\cos x_4$	.5135955	.1503312	3.42	0.001	.2189519	.8082392
$\sin x_4$	-.3552356	.1097121	-3.24	0.001	-.5702674	-.1402038
$\cos x_5$	.0824003	.0590216	1.40	0.163	-.0332799	.1980804
$\sin x_5$	.1111981	.0258736	4.30	0.000	.0604868	.1619094
EU	.0698374	.0327915	2.13	0.033	.0055673	.1341076
$EURO$	.0061857	.0082689	0.75	0.454	-.010021	.0223924
GDP	-.0031762	.003979	-0.80	0.425	-.0109748	.0046225
π_k	-.0068753	.0074374	-0.92	0.355	-.0214523	.0077017
r_k	.0051477	.0090592	0.57	0.570	-.0126081	.0229034
$M2_k$	.000221	.0000676	3.27	0.001	.0000884	.0003535
$HERF_k$	-.0160632	.047639	-0.34	0.736	-.109434	.0773076
$C.CR_k$	6.903923	1.639471	4.21	0.000	3.69062	10.11723
$C.TR_k$	.5577072	.1561363	-3.57	0.000	-.8637287	-.2516858
$C.LR_k$	.0039784	.000822	4.84	0.000	.0023672	.0055895
$C.InsR_k$	-.0971382	.109286	-0.89	0.374	-.3113349	.1170584

The estimation outcome shows that the (LL) function converges at the value of 390.46699 after 26 iterations. The model is estimated using SFA approach with a half-normal distribution specified for the model's inefficiency term.

For the purpose of interpreting the frontier variables' estimation results, it is assumed that all observed banks pursue the objective of cost minimization. Accordingly, a cost frontier is defined as the "minimum expenditure required to produce any scalar output, given input prices" (Kumbhakar and Lovell, 2000, p 33 – 34), implying that the expenditure of each observed bank must be on or above the cost frontier. This relationship is depicted by Figure 9 below. Time variables in the stochastic frontier model R-F-C defined above account for technological changes that are detected by the frontier's shifts. Accordingly, time-related shifts in the cost frontier are caused by the significant impact (estimated coefficient) of any time or time-interactive variables describing the frontier function.

The frontier can be depicted according to Battese and Coelli (1995, p 329) as follows:

Figure 9: **A Cost Frontier (Kumbhakar and Lovell, 2000, p 33)**



The frontier shown above is a single-dimensional cost frontier used for the sole purpose of illustrating the impact of the cost frontier's shift. It is a simple representation of the true frontier which would be shaped as a surface since total costs are regressed against several independent variables including outputs, input prices, time interactive variables, Fourier terms etc. The figure above hows a downward shift in the cost frontier over time which translates into enhanced cost efficiency. This is because for a given level of outputs, a bank would incur less costs to produce the given level of outputs, or that for the same level of costs more outputs can be produced. In both cases the result is better usage of resources that is reflected in better cost efficiency (lower costs).

Considering the estimation results above, it appears that there are 3-time interactive terms showing a significant impact at the 5% critical level on total costs over 2010 – 2018 but with different directions. tlw_2 (representing the interaction of time with the price of funds) is found to cause an upward shift in the cost frontier causing cost efficiency to decline over time. Similarly, tly_3 (representing the time interaction with the third output, off-balance-sheet items OBS) is also found to drive the cost frontier upwards. On the other hand, tly_2 (representing the time interaction with the second output, other earning assets or OEA) is showing a negative impact suggesting that there have been favorable technological advances related to producing OEA which seems to cause a downward shift in the cost frontier (i.e., enhanced cost efficiency). Estimates for the latter time-interactive factor (tly_2) show that it has the most significant negative coefficient in terms of size and statistical significance.

All in all, the collective impact of these three significant time-interactive terms seems to suggest that the cost frontier follows a non-linear trend over time. The cost frontier function can reveal a maximum stationary point(s) relating total costs and time: this can be seen by twice differentiating it with respect to time (given the negative sing of the 2nd order time interactive term, $\frac{1}{2}t^2$).

More generally, the cost function estimated above shows that 56% of the frontier's variables (27 out of 48) are found to be significant at 5% (45% of these variables are significant at 1%, that is 22 variables). The model's Translog terms (including time-interactive terms and time dummies) seem to perform well in explaining the variation of total costs as the majority of them display significant coefficient estimates. Specifically, 15 out of the 27 Translog terms (i.e. 55% of these terms) are significant at 5% critical level, indicating a good fit to the data by Translog terms. Compared to published research, the Translog terms in the R-F-C model seem to perform well, as

Fitzpatrick and McQuinn (2008) for instance using Translog functional form, found 44% only of the frontier variables as significant at 5% with an additional 1 variable found significant at 1%.

As for EU-membership and Euro-adoption dummies (EU and $EURO$), results show that only the EU factor has a positive and significant relationship with total costs at 5% critical level. This result indicates that, over 2010 – 2018, banking systems of Member States (i.e. EU-economy) would incur higher costs than banks operating in a non-EU-economy. Specifically, estimation results indicate that banks operating in an EU-economy are likely to run higher costs by around[67] 7% (.0698374) compared to banks operating in a Non-EU-economy. That is the significant impact of EU is found to cause banking systems of Member States to incur around 7% higher costs than banking systems of Non-EU countries.

Concerning the impact of the macroeconomic variables, it appears that $M2_k$ has a significant and positive impact on total costs. $M2_k$ is the ratio of M2 money supply to GDP that is used as a proxy for the level of bank intermediation in economy k. Therefore, the greater the value of $M2_k$, the greater the degree of banks' intermediation (or financial depth) in the financial system. Results indicate that, in a given economy, banks' costs are likely to increase as they become more involved in financial intermediation (taking deposits and making loans), however such increase is minor given the relatively small size of $M2_k$'s coefficient (.000221). Despite that this result indicates that greater financial depth suggests more costs incurred, this does not imply lower cost efficiency for a given banking system as costs could increase less proportionately than outputs. Previous studies have tested the correlation between financial intermediation and cost efficiency but not total costs. Kasman and Yildirim (2006) use SFA in a single-stage estimation approach to estimate cost and profit functions (truncated at the 2nd order terms) for commercial banks in 8 new member

[67] If both the dependent and independent variables are logged, then the interpretation of coefficient estimates (β) follows the elasticity interpretation, that is, when x increases by 1%, y increases by β%. On the other hand, if the dependent is logged and the independent is not logged, then the interpretation of coefficient estimates (β) follows the percentage interpretation: when x increases by 1%, y increases by β*100% (Wooldridge, 2003).

states of the EU in 2004. They found that cost efficiency is positively correlated with the level of financial development (M2/GDP).

In relation to the country-level risk variables, three out of the four risk factors introduced into the frontier functional form seem to have very significant impact on total costs at the 1% critical level. For country-specific credit risk $C.CR_k$, calculated as the country-averaged ratio of loan loss provisions to total loans, estimation shows that it has a positive and significant impact on European banks' total costs. The positive correlation of $C.CR_k$ with total costs indicates that as the proportion of loan provisions to the size of loans portfolio increases by 1%, total cost would tend to increase by 6.9% (given the coefficient of $C.CR_k$ of 6.903923). In other words, for an average increase of 1% in a given banking system's $C.CR_k$, total costs of that country's banking system are likely to increase by around 6.9% on average. Notably, $C.CR_k$ is found to have the most significant impact in magnitude on costs amongst the other risk factors, which can indicate that loan loss provisions are one major factor affecting European banking costs.

As for country-specific trading risk $C.TR_k$, calculated as the country-averaged ratio of net trading revenue to total revenues, results show a significant and positive impact of $C.TR_k$ on total costs. Specifically, for a 1% increase in the value of net trading income relative to total income, total costs are likely to increase by nearly 5.5% (coefficient of .5577072). The positive relationship between an increase in net trading revenue (the numerator effect) and total costs can only be explained as follows. The increase in net trading revenue (the numerator) can be the result of a more proportionate increase in trading revenues relative to the increase in trading costs (losses), leading to larger net trading revenue as a result. Accordingly, the positive correlation between net trading revenue and total costs can only be explained by a parallel, but less proportionate, increase in trading costs that is translating into increasing losses. Therefore, this positive correlation between an increase in net trading revenue (hence increasing the value of $C.TR_k$) and total costs is probably the result of increasing trading costs as trading activities expand on banking system level. Accordingly, data over 2008 to 2018 seems to suggest that the European banking systems seem to incur increasing costs as they expand their trading activities.

Finally, results for country-averaged liquidity risk $C.LR_k$ show a positive and significant impact on total costs with a relatively small magnitude of (.0039784). $C.LR_k$ is calculated as the ratio of customer and short-term funding to liquid assets. Results show that for a 1% increase in liquidity risk (or in customer and short-term funding relative to liquid assets), total costs increase by about 0.004%. It is worth noting that the combination of customer and short-term funding can suggest a funding-mix effect on total costs, however to more closely represent the lines of defense banks use against short-term liquidity shortages, both sources of funding had to be included to more realistically represent banks' exposure to liquidity risk. The funding-mix effect can be isolated by comparing the cost of customer and short-term funding.

All in all, estimation results on banking system level indicate that, over 2010 – 2018, total costs for European banking systems tend to increase relative to the depth of the financial system. Moreover, banking systems of Member States tend to incur higher costs than banks operating in non-EU-economies. Lastly, results show a positive correlation between total costs, loan losses, exposure to trading risk and money market (or short term) funding.

5.4.5 Determinants of Technical Cost Inefficiency

This section discusses the determinants of technical cost inefficiency model defined in (19). The estimates of these inefficiency correlates are produced simultaneously with those of the cost frontier R-F-C. As pointed earlier, Battese and Coelli (1995) time-flexible technical inefficiency model is followed in this research where the technical inefficiency is regressed against a set of potential explanatory variables (determinants or correlates). These determinants of cost inefficiency are bank-specific variables that can be classified into 7 categories. Full description of these variables is presented in Table 3. These variables include:

1. **Risk Factors** that include: (a) an ex ante measure for credit risk represented

by the credit risk appetite (CRA) which is calculated as the ratio of risk weighted assets to total assets, (b) an ex post measure for credit risk represented by the ratio of loan loss provisions to total loans ($CrdtRsk$), (c) a proxy for trading risk ($TrdgRsk$)) calculated as the ratio of net trading revenue to total revenues, (d) a proxy for liquidity risk ($LqdtyRsk$) calculated as the ratio of customer and short-term funding to total funding, and (e) a proxy for insolvency risk ($InslvcyRsk$) calculated as the ratio of equity capital to total assets.

2. **Business Structure** (OBS) represented by the ratio of the credit equivalent of Off-Balance-Sheet items (OBS) to total assets.
3. **Income Structure** ($NetInt$) represented by the ratio of interest income to total income.
4. **Funding Structure** ($Funding$) represented by the ratio of short-term funding customer deposits to total funding.
5. **Bank Size** represented by a non-linear form of assets ($Assets$) following Berger et al (1993)[68].
6. Two forms of **time dummies** t and $\frac{1}{2}t^2$ re included to ensure the time-flexibility of the technical inefficiency model so as to capture the change in inefficiency effects over time[69].

Estimation results are shown in Table 17 below:

[68] The idea of applying the second-order logarithmic form of (assets) stems from the work of Berger and Mester (1997, p 924) who apply a second-order logarithmic form of credit risk measure (NPL/Loans) as a control variable in the cost function to control for the effect of bad luck. The second-order logarithmic form of (assets) is applied because the model could not converge with other forms of (assets), and including a linear (log of assets) along with this non-linear form of (assets) has demonstrated considerable multicollinearity.

[69] Empirically, it is the mean or the variance of the inefficiency term's distribution that is regressed conditionally against the proposed set of determinants or z variables (according to the inefficiency models suggested by Battese and Collie 1995 and Kumbhakar and Lovell 2000 respectively). The research applies the conditional variance model of Kumbhakar and Lovell (2000) as the conditional mean model was tested but convergence could not be achieved under which. However, results are interpreted as the inefficiency term being correlated with a set of z variables (see for instance Girardone, Molyneux and Gardener, 2004). Neither the mean nor the variance of the inefficiency term's distribution should be thought of as constants because every observation's inefficiency term will have a unique distribution with different mean or variance as the sample is drawn from an unknown population. The conditional variance model is also consistent with the half-normal distributional assumption of the inefficiency term since for the conditional mean model half-normality cannot be assumed since the mean would be zero. The latter can accommodate the truncated normal distribution with truncation at zero.

Table 17: Determinants of Technical Cost Inefficiency[70]

u_{it}	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
CRA	.2328846	.9673399	0.24	0.810	-1.663067	2.128836
$CrdtRsk$	-25.53426	23.98459	-1.06	0.287	-72.5432	21.47468
$TrdgRsk$	**8.940631**	3.473104	2.57	**0.010**	2.133472	15.74779
$LqdtyRsk$	-.0130596	.0132164	-0.99	0.323	-.0389633	.0128441
$InslvcyRsk$	**-15.67597**	6.46373	-2.43	**0.015**	-28.34465	-3.007292
OBS	.3149656	.5511261	0.57	0.568	-.7652219	1.395153
$NetInt$	1.063079	2.209279	0.48	0.630	-3.267028	5.393541
$Funding$	2.229046	1.523265	1.46	0.143	-.7564979	5.214589
$\frac{1}{2}\ln(Assets)^2$	-.01226	.0078235	-1.57	0.117	-.0275938	.0030737
t	-.3569105	.5751851	-0.62	0.535	-1.484253	.7704315
$\frac{1}{2}t^2$	.1234843	.1954843	0.63	0.528	-.259658	.5066265

Prior to discussing the results, it should be noted that bank-specific variables described above are correlated with inefficiency effects as defined in (55). Therefore, a variable that is shown to have a significant and negative coefficient (such as $InslvcyRsk$) suggests that it is negatively correlated with cost inefficiency. This can alternatively be interpreted as $InslvcyRsk$ being positively correlated with cost efficiency. Estimates highlighted with bold typeface are the statistically significant factors at the 5% critical level.

Results show that cost inefficiency seems to significantly be influenced by two main factors: trading risk ($TrdgRsk$) and insolvency risk ($InslvcyRsk$). Trading risk ($TrdgRsk$), represented by the ratio of net trading revenue to total revenues, is found to have a positive and significant impact on cost inefficiency. Taking on more trading risk, i.e. the larger the value of ($TrdgRsk$), results either from larger numerator (net trading revenue) or smaller denominator (total revenue). Moreover, the larger value of ($TrdgRsk$) can reflect the business-mix effect as in this case the bank chooses to generate more income from trading as opposed to other activities related to the

[70] Statistically-significant term are shown in **bold**.

banking book for instance. Both scenarios suggest greater involvement in trading. Therefore, the finding suggests that as banks take more trading risk, i.e. as net trading revenue increases relative to total revenues, European banks become less cost efficient. A negative impact on cost efficiency can possibly be explained by increasing trading losses incurred by European banks on average over 2010 – 2018.

On the other hand, insolvency risk ($InslvcyRsk$), calculated as the ratio of equity capital to total assets, is found to have a negative impact on cost inefficiency. In other words, banks are likely to be more cost efficient if they are better capitalized. This finding is consistent with that of Mester (1993 and 1996) and Girardone et al (2004) who also find a negative correlation between capital ratio (equity/assets) and cost inefficiency. Likewise, Pasiouras (2008) also applies the capital ratio and finds a positive correlation with technical efficiency using Greek banking data. This is also consistent with the findings of Isik and Hassan (2003) using data on Turkish banking, Rao (2005) using data on United Arab Emirates and Kwan and Eisenbeis (1997) using US data. There are four potential interpretations to this positive correlation between capitalization and efficiency.

First, well-capitalized banks are more capable of expanding their balance sheets and further spread there costs over higher levels of outputs which would translate into reduced average costs and ultimately enhance the bank's cost efficiency. Second, the positive relationship between $InslvcyRsk$ and cost efficiency can be explained from the cost of finance perspective. Better capitalization is associated with lower levels of default risk which would be perceived by regulators, investors, and capital markets providing long term finance as a good sign of the bank's safety. As a result, regulators would require less regulatory capital to be set aside, investors would ask for less expected returns to compensate for the risk they are taking, and capital markets would translate better capitalization into less interest required on long term finance. This is based on the moral hazard argument of Mester (1996) such that, better capitalization reduces the potential for moral hazard by bank mangers which in turn would positively be reflected on the short- as well as long-term cost of finance. In any of these cases, better capitalized banks can enjoy lower long-term funding costs and therefore become more cost efficient.

Third, an alternative view based on the argument of Girardone et al (2004) suggest that, the inverse relationship between capital and cost inefficiency can be explained as: banks with low cost inefficiency are more capable of generating profits and therefore retaining more earnings as capital. Fourth, greater capitalization is believed to reduce the adverse impact of principal-agent problem stemming from information asymmetry. In this sense, Pasiouras (2008) argues that the level of equity capital reflects shareholders incentive to monitor the bank management to ensure that it operates efficiently since equity is their capital that is at risk. Hence greater capitalization in this sense would result in greater technical efficiency.

Notably, insolvency risk ($InslvcyRsk$) displays the largest and most significant coefficient amongst all other bank-specific cost efficiency determinants. This strong relationship indicates that bank capitalization is the main factor affecting its technical cost efficiency. With this in mind, the conclusion therefore is that: European banks can operate more cost efficiently if they are better capitalized. It should be noted here that this conclusion should be treated with caution, as it should not be interpreted as increasing the ratio of equity funding to total assets (as represented by $InslvcyRsk$) will improve cost efficiency at any level. This is because the right balance of equity to debt finance needs to be established as a bank will surely be cost inefficient if it is excessively financed by equity since equity is more expensive source of finance than debt. On the other hand, equity finance is of a unique importance for banks as it delivers safety that distance the bank from the risk of insolvency (which is paramount not only for individual banks but also for the banking systems they are part of), which is why the argument of moral hazard for instance applies. The impact of bank safety on cost efficiency is what the insolvency risk measure ($InslvcyRsk$) attempts to capture. In this sense, better capitalized banks can improve their cost efficiency as results suggest but up to the level where equity capitalization achieves the right balance between the banks' risk profile, insolvency, and cost of finance. Beyond this point, further equity capitalization would cause costs to increase dramatically and therefore increase cost inefficiency as a result.

Turning to a different issue, estimation results seem to reveal more interesting findings. The credit risk appetite ratio (CRA) is found to have no significant impact on European banks cost efficiency. The size of the CRA ratio indicates to the level of exposure to credit risk, and the impact of CRA on cost efficiency indicate to how well-managed credit risk is. Finding no significant impact of CRA is quite interesting as it could be an indication of the ineffectiveness of the capital requirements set by Basel I as far as the criteria of risk-weighted assets is concerned. Having found no significant impact by CRA on cost efficiency provides no indication on how well credit risk is managed by European banks over the sample period 2008 – 2018. If CRA was effective in representing credit risk from an ex ante perspective, it could have shown a significant positive or negative impact on banks cost efficiency.

In this sense, if CRA had a positive and significant impact on bank cost efficiency, this would have indicated that credit risk is well managed by European banks. This suggests that taking on more credit risk is having a positive impact on costs (in terms of lesser defaults) as loans mature which could be a direct result of the diversification effect. On the other hand, if CRA was to show a negative and significant impact on cost efficiency, this would have reflected an excessive credit risk that is being reflected in increasing defaults. However, having found no significant impact for CRA on cost efficiency may well indicate to the insensitivity of Basel I regulatory risk weights since they are based on the criteria of one-size-fits-all (Dowd, 2002).

Moreover, the ratio of interest margin to total revenues (as represented by $NetInt$) is showing no significant impact on European bank's cost efficiency. This is in contrast to the findings of Girardone et al (2004) where the ratio of interest margin to total assets was found with a negative impact on cost efficiency in Italian banks over 1993 – 1996.

Furthermore, cost efficiency correlates show no significant impact of ex post credit risk $CrdtRsk$ (represented by the ratio of loan losses to total loans) on cost efficiency. Estimation result for $CrdtRsk$ is quite mixed as credit risk is found to be negatively correlated with cost inefficiency, which is counterintuitive, however this is not statistically supported given the insignificance of this estimated coefficient. What is more, the size of OBS items relative to the bank total assets as represented by OBS factor also shows no significant impact on cost efficiency.

Lastly, concerning the relationship between size, represented $\frac{1}{2}\ln(Assets)^2$, and cost efficiency, results show no evidence of a significant impact of size on cost efficiency. This is further confirmed by conducting a correlation (covariance) test between bank size (total assets) and cost efficiency scores. The test reveals very weak correlation (a coefficient of -0.0372) that is insignificant at 5% or 10% level (p-value = 0.6134). Such finding is in line with Mester (1993), Berger and DeYong (1997), Altunbas et al (2000) and Girardone et al (2004). The conclusion therefore is that there is no strong evidence to suggest that larger banks are more cost efficient than smaller banks or vice-a-versa. These results are in contrast to, for instance, Cavallo and Rossi (2001) who apply the two-stage estimation technique and specify a Translog cost function and the time-inflexible inefficiency model, Battese and Coelli (1992), to banks in 6 major EU countries over 1992 – 1997 and find that smaller banks are more cost efficient than larger banks.

To sum up, estimation results for bank-specific cost efficiency determinates over 2008 – 2018 seem to suggest the following: European banks can operate more cost efficiently by (1) reducing their exposure to trading risk (in terms of reducing trading losses), and (2) being better capitalized. Moreover, it seems that Basel I credit risk weighting criteria has no significant impact on European banks cost efficiency given the insignificance of the credit risk appetite measure (CRA) found. Further, there seems to be no significant correlation between cost efficiency and the level interest margin (relative to total revenues) or the level of OBS items (relative to total assets). Finally, there is also no evidence to suggest that cost efficiency is related to bank size.

5.5 Empirical Results 2: Profit Function (R-F-P)

Prior to conducting the estimation on the preferred profit model R-F-P and produce the corresponding profit efficiency estimates, there are three main tests that need to be applied to the model, these, according to Battese and Coelli (1995, p 330), are: technical inefficiency test, stochasticity test, and inefficiency correlates test. The aim of these tests is to ensure that the profit model (1) has an inefficiency effect, (2) it is stochastic, and (3) if the proposed inefficiency term's correlates are jointly significant.

5.5.1 Hypothesis Testing

5.5.1.1 Technical Profit Inefficiency Test

This test investigates whether the cost model R-F-P has inefficiency effects or not. This entails testing $H0:\sigma_u^2 = 0$ against the alternative $H1:\sigma_u^2 > 0$ since inefficiencies can only be nonnegative. The results of this test are resented in the table below:

Table 18: Technical Profit Inefficiency

Components of the composite error term	Coeff.	Std. Err.	z	P|z|
Variance of the Random Error component ($\ln\sigma_v^2$)	-25.61103	20.04917	-1.28	0.201
Variance of the Technical Inefficiency component ($\ln\sigma_u^2$)	-4.324954	.1034222	-41.82	0.000

Findings clearly show that the null hypothesis of absent inefficiency effects is strongly rejected since $\ln\sigma_u^2$ is found to be statistically very significant at the 1% level. Hence the R-F-P profit frontier model has a technical inefficiency effect.

5.5.2 Stochasticity Test

Having ensured the existence of the inefficiency effects, these effects will need to be checked for stochasticity. Stochasticity test involves testing the significance of the parameter $\gamma = \sigma_u^2/\sigma^2$ which in effect tests if the variation of the inefficiency term is statistically significant in relation to the total variation of the composite error term.

Accordingly, the null states that the technical inefficiency is not stochastic, that is

$H_0 : \gamma = 0$ implying that frontier is not stochastic, and the alternative $H_1 : \gamma > 0$ (as technical inefficiency is always assumed as being nonnegative) implies that the frontier is stochastic. Not rejecting the null suggests that the frontier's parameters can consistently be estimated using ordinary least squares OLS where the inefficiency term can be removed from the model (Coelli, 1996, p 5). Results show that the null hypothesis is comfortably rejected at 1% critical level: LR statistics ($\bar{\chi}^2_{01}$) = 1.1e+02 with P-value = 0.000. Therefore, the variation of the technical inefficiency term is significant in relation to the variation of the composite error term hence the profit model R-F-P is stochastic, implying that the model should be estimated according to the Stochastic Frontier Analysis (SFA) approach.

5.5.3 Inefficiency Correlates Test

This test tests the null hypothesis that the inefficiency term's correlates are jointly insignificantly different from zero, although the individual effects of one or more of the correlates (determinants) may be statistically insignificant. Given the inefficiency model specified in equation ((59)(19)), $H_0 : \beta_1 = \beta_2 = ... = \beta_{11} = 0$. The test statistics is Chi-squared distributed with its parameter being equal to the parameters assumed to be zero. Results show that H_0 is comfortably rejected at the 1% critical level where the tests statistics chi2 (11) = 489.13 and Prob. > chi2 = 0.0000.

All in all, the results of the above three tests indicate that the preferred profit frontier model R-F-P has a technical inefficiency effects and the model is stochastic, and that the suggested correlates of the profit inefficiency term are jointly significant. Having found that, the next step now is to estimate the R-F-P model as defined in equation ((55)) simultaneously with the inefficiency model defined in equation ((19)) according to the single-stage SFA approach.

5.5.4 Estimating Profit Function (R-F-P)

Crucial to producing efficient parameters estimates (lowest standard errors possible) is to ensure convergence of the estimated function. Convergence is important to achieve since it indicates that the greatest value of the log-likelihood function (LL) which corresponds to the R-F-P model is achieved. Without achieving the convergence of (LL), the resulting parameters' estimates would be less reliable (or less efficient). In light of this, Figure 10 graphically displays the convergence of the (LL) function representing the profit frontier model which comprises of the frontier model R-F-P defined in ((55)) and the corresponding inefficiency model defined in ((59)(19)).

Figure 10: Convergence of R-F-P frontier model with Battese & Coelli (1995) inefficiency model and half-normal distributional assumption specified

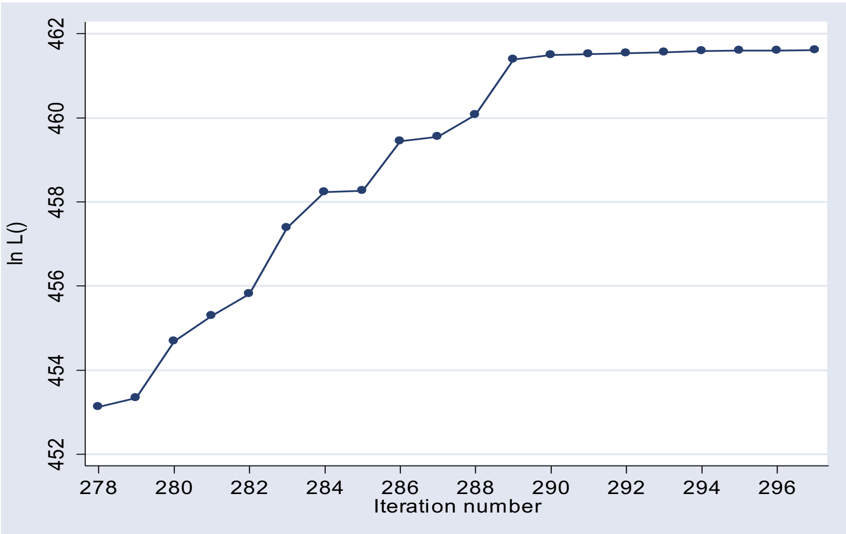


Figure 10 depicts the progression of the iterative process applied by the maximum likelihood estimation technique MLE to evaluate the LL function. Convergence is achieved when the curve becomes eventually flat. Figure 6 shows that the profit function converges to its maximum possible value of 460.06731 after 297 iterations. The estimation outcome for the profit frontier model is shown below:

Iteration 295: log likelihood = 459.43983

Iteration 296: log likelihood = 459.55414

Iteration 297: log likelihood = 460.06731

Stoc. frontier normal/half-normal model

Number of obs = 541

Wald chi2(48) = 5003258.07

Prob > chi2 = 0.0000

Log likelihood = 460.06731

$\ln P_{it}$	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
lw_1	-.1357753	.0589063	-2.30	0.021	-.2512296	-.0203211
lw_2	-.4223962	.0597759	-7.07	0.000	-.5395548	-.3052375
ly_1	1.619623	.4163719	3.89	0.000	.803549	2.435697
ly_2	-.2658352	.2012139	-1.32	0.186	-.6602071	.1285367
ly_3	.0005074	.0522326	0.01	0.992	-.1018667	.1028814
$\frac{1}{2} lw_1 lw_1$	.2841759	.034546	8.23	0.000	.216467	.3518848
$\frac{1}{2} lw_2 lw_2$	.0722887	.0089373	8.09	0.000	.054772	.0898054
$lw_1 lw_2$	-.0359784	.0035643	-10.09	0.000	-.0429643	-.0289924
$\frac{1}{2} ly_1 ly_1$	-.8424527	.1907209	-4.42	0.000	-1.216259	-.4686466
$\frac{1}{2} ly_2 ly_2$	.1688598	.0873803	1.93	0.053	-.0024024	.340122
$\frac{1}{2} ly_3 ly_3$	-.0349755	.0244934	-1.43	0.153	-.0829817	.0130306
$ly_1 ly_2$	-.0746498	.0106057	-7.04	0.000	-.0954365	-.053863
$ly_1 ly_3$	.0450624	.0110283	4.09	0.000	.0234474	.0666775
$ly_2 ly_3$	.0050695	.0053627	0.95	0.344	-.0054412	.0155803
$lw_1 ly_1$	.0475205	.0112738	4.22	0.000	.0254242	.0696168
$lw_1 ly_2$	-.0062002	.0090865	-0.68	0.495	-.0240094	.0116091
$lw_1 ly_3$	.0119354	.007757	1.54	0.124	-.0032681	.0271389
$lw_2 ly_1$	.0184969	.0034657	5.34	0.000	.0117044	.0252895
$lw_2 ly_2$	.0105026	.0046663	2.25	0.024	.0013568	.0196485
$lw_2 ly_3$	-.0123641	.00287	-4.31	0.000	-.0179893	-.006739
tlw_1	.0094042	.0037032	2.54	0.011	.0021462	.0166623
tlw_2	-.0032676	.0007813	-4.18	0.000	-.0047989	-.0017363
tly_1	.0123355	.0035433	3.48	0.000	.0053907	.0192803
tly_2	.0049389	.0033683	1.47	0.143	-.0016629	.0115406
tly_3	.0001776	.0016585	0.11	0.915	-.003073	.0034282
t	-.0656819	.0162856	-4.03	0.000	-.0976011	-.0337627
$\frac{1}{2} t^2$	.0104154	.0029972	3.48	0.001	.004541	.0162898
$\cos x_1$	.2418113	.0533592	4.53	0.000	.1372291	.3463935

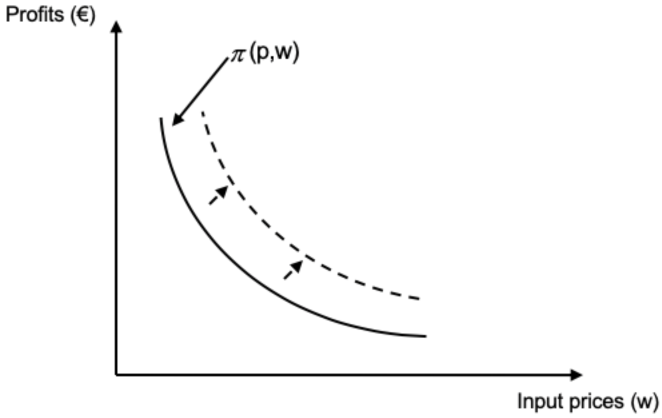
$\sin x_1$	.124142	.0438299	2.83	0.005	.038237	.210047
$\cos x_2$	-.1038874	.0546648	-1.90	0.057	-.2110285	.0032537
$\sin x_2$	.0209514	.0297249	0.70	0.481	-.0373083	.0792111
$\cos x_3$	.0098205	.0208366	0.47	0.637	-.0310185	.0506596
$\sin x_3$	-.0054497	.0086593	-0.63	0.529	-.0224216	.0115222
$\cos x_4$	-.2264442	.0735719	-3.08	0.002	-.3706424	-.082246
$\sin x_4$	.311149	.0233132	13.35	0.000	.2654561	.3568419
$\cos x_5$	-.175989	.0262042	-6.72	0.000	-.2273483	-.1246297
$\sin x_5$	-.1075219	.0204796	-5.25	0.000	-.1476611	-.0673826
EU	.0397273	.0125506	3.17	0.002	.0151286	.0643261
$EURO$	-.0455074	.0065705	-6.93	0.000	-.0583852	-.0326295
GDP	-.0023406	.0019989	-1.17	0.242	-.0062584	.0015772
π_k	.0082125	.0034989	2.35	0.019	.0013548	.0150703
r_k	-.011919	.0024669	-4.83	0.000	-.0167541	-.0070839
$M2_k$	.0000542	.0000397	1.37	0.172	-.0000236	.000132
$HERF_k$	.0040224	.0242246	0.17	0.868	-.0434569	.0515017
$C.CR_k$	-.4141348	.4192986	-0.99	0.323	-1.235945	.4076754
$C.TR_k$	-.1556754	.1133088	-1.37	0.169	-.3777565	.0664057
$C.LR_k$	-.0013971	.0003116	-4.48	0.000	-.0020078	-.0007864
$C.InsR_k$	-.0099405	.0519572	-0.19	0.848	-.1117748	.0918939

The estimation outcome shows that the (LL) function converges at the value of 460.06731 after 297 iterations. The model is estimated using SFA approach with a half-normal distribution specified for the model's inefficiency term. The Wald Chi-squared test confirms that the model's 48 parameters (representing the frontier variables) are jointly very significant at the 1% level in explaining the variation of the dependent variable, total profits ($\ln P_{it}$).

For the purpose of interpreting the frontier variables' estimation results, it is assumed that all observed banks pursue the objective of profit maximization. Accordingly, a profit frontier corresponds to the maximum profits attainable from producing a given output level and mix under given input prices (Kumbhakar and Lovell, 2000). The profit frontier represents the best performance possible and therefore is applied as a bench mark against which the observed profits of sample banks are compared. As the maximum obtainable profit is characterized by efficient profit frontier, observed banks' profits must lie on or beneath the frontier, causing profit efficiency scores to range between (0) and (1).

With regard to the time and time-interactive terms, estimation results show that 5 out of a total of 7 variables are found to significantly affect the profit frontier R-F-P. As illustrated earlier, any significant impact of these time interactive terms cause the frontier to shift upwards or downwards depending on the sign of their estimated coefficients. The inclusion of these time and time-interactive terms as frontier variables serves the purpose of detecting the impact of technological changes over time (Battese and Coelli, 1995). Figure 11 below is introduced to simply illustrate the implications of shifts in the profit frontier. Figure 7 shows an upward frontier shift that translates into higher profits resulting from the same level of expenses incurred which suggests enhanced profit efficiency.

Figure 11: **A Profit Frontier (Kumbhakar and Lovell, 2000, p 38)**



Estimation results show that the time-interactive terms tlw_1, tlw_2, and tly_1 are all significant at 1% critical level.

The time-price of funds interactive term with the price of funds (tlw_1) is found with a positive and significant coefficient resulting in an upward shift in the profit frontier, this is due to favorable changes over time in the price of funds that are causing profits to increase. Likewise, estimation shows that the time-loans interactive term (tly_1) has a positive and significant impact suggesting that technological changes in producing outputs seem to also contribute to an upward shift in the profit frontier for European banks, leading to higher levels of profits as a result. This time-interactive factor is found with the largest significant and positive coefficient amongst the other two significant time-interactive terms.

On the other hand, the third term which involves time interaction with the price of real capital (tlw_2) is found with a significant and negative coefficient causing a downward shift in the profit frontier over time due to changes in the price of real capital. The overall impact of these three significant time-interactive terms seems to suggest that the profit frontier follows a non-linear trend over 2010 - 2018. The profit frontier function R-F-P can reveal a minimum stationary point(s) relating profits and time (this can be seen by twice differentiating it with respect to time given the positive sing of the 2nd order time interactive term, $\frac{1}{2}t^2$).

With regards to the EU-membership and Euro-adoption factors, EU and $EURO$, estimation shows that both are found to significantly affect European banks' profits over 2008 – 2018 at the 1% critical level. EU-membership factor is showing a positive impact of (.0397273) while Euro-adoption factor is showing a negative impact of (-.0455074). These findings indicate that operating in the EU Member States has positively influenced profits for banks operating in these countries however adopting the euro by some countries seems to negatively affect profits. Specifically, results seem to suggest that operating in an EU- but Non-euro-economy (such as the UK and Sweden) provides a better economic environment for banks to make more profits.

Part of this conclusion might seem in contradiction to the impact of the EU membership factor (EU as a frontier variable) on total costs where being a Member State was found to cause bank total costs to increase. Nonetheless, incurring higher costs does not suggest that profits will be lower, as profits and costs can both increase disproportionately. Therefore, the only way that the estimated impact of EU on profits and total costs can make sense is that: for EU banking systems, profits and costs can increase simultaneously, but disproportionately. More specifically, estimation results show that costs tend to increase more proportionately than profits on a banking system level given the estimated impact (i.e. magnitude or coefficient) of the EU factor on profits and total costs under the profit and cost functions which is around 0.4 and 0.7 respectively.

The estimation outcome also shows that there are two macroeconomic factors with significant impact on European banking profits over 2008 - 2018, these are: inflation rate (π_k) and short-term interest rates (r_k). Results show that profits are positively associated with the inflation rate. This reiterates the fact that the banking business is pro-cyclical as banks tend to do well in good times where economic growth is positive. The inflation rate under these conditions tends to be higher which explains the positive correlation found between inflation rate and bank profits.

On the other hand, short-term interest rates are found to negatively and significantly affect banks' profits (-.011919). That is, for every 1% increase in short-term interest rates, European banks profits tend to decline by around 1.19% on average over 2010 – 2018. Such result is economically justified as increasing short-term interest rates tightens the bank's interest rate spread since commercial banks are exposed to the yield-curve effect because they borrow short and lend long. Accordingly, if short-term interest rates become lower (higher) than long-term rates suggesting positive (negative) slope of the yield curve and implying looser (tighter) monetary policy, banks can enjoy wider (tighter) interest rate spreads which can ultimately and positively (negatively) affect profits.

With regards to country-specific (or banking systems-specific) risks, estimation outcome shows that the only risk factor with significant impact on European banks' profits is country-specific liquidity risk $C.LR_k$. Country-specific liquidity risk $C.LR_k$, calculated as the country-averaged ratio of customer and short-term funding to liquid assets, is found with a significant and negative correlation with profits (-.0013971). This finding suggests that the higher the liquidity risk or the higher the value of $C.LR_k$, the lower the profitability will be.

Increasing liquidity risk can be the result of the denominator effect (lower level of liquid assets), or numerator effect (increasing level of customer deposits and money market funding relative to liquid assets holdings).

The only logical way to interpret this result is by considering the numerator effect, hence this result suggests that relying more on customer and short-term funding is negatively affecting European banks' profits. However, this interpretation embodies the funding-mix effect so it is important to isolate or decompose the funding-mix to reach a more accurate conclusion. Results presented on a bank-specific level in the following section (Table 19) clearly shows that increased reliance on customer deposits relative to total funding is found to have a significant positive impact on profit efficiency. Accordingly, it can be inferred that the money market funding part of the funding mix in the numerator effect is what is causing the negative correlation between liquidity risk and profits. This is also consistent with the earlier finding in that liquidity risk was found as positively correlated with total cost such that: the greater the exposure to liquidity risk in the sense of using more short-term market funding, having isolated the funding-mix effect, the higher the costs. Therefore, the overall conclusion that can be drawn given results above is that: greater reliance on money market funding significantly increases total costs and reduces profits at a banking system level.

All in all, estimation results at banking system level seem to suggest that profits tend to increase if banks operate in an EU- but non-euro economy. For EU banking systems, results show that profits and costs can increase simultaneously, however, with the greater impact of the EU factor on total costs in mind, can increase more proportionately than profits given the estimated impact (i.e. magnitude or coefficient) of the EU factor on profits and total costs. Further, results show that profits are positively associated with the inflation rate and negatively correlated with short-term interest rates. Finally, results seem to suggest that the lower the liquidity risk, in the sense of using less money market funding, the higher the profitability in European banking.

5.5.5 Determinants of Technical Profit Inefficiency

In this section, the determinants of technical profit inefficiency will be discussed as defined by the inefficiency model. These estimates are produced simultaneously with those of the profit frontier R-F-P defined above. The inefficiency model follows is Battese and Coelli (1995) time-flexible technical model specification assuming half-normal distribution for the inefficiency term. Technical profit inefficiency term is regressed against the same bank-specific determinants used to explain the variation of technical cost inefficiency so as to draw consistent conclusions and to investigate any possible correlation between bank-specific profit and cost efficiencies. Table 19below shows the estimation results for the different categories of bank-specific variables:

Table 19: Determinants of Profit Inefficiency

u_{it}	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
CRA	-.2489303	.6231717	-0.40	0.690	-1.470324	.9724638
$CrdtRsk$	**135.9684**	25.42431	5.35	0.000	86.13771	185.7992
$TrdgRsk$	**-12.2112**	4.101109	-2.98	0.003	-20.24922	-4.173174
$LqdtyRsk$	**-.0212271**	.0083284	-2.55	0.011	-.0375506	-.0049037
$InslvcyRsk$	-5.011016	6.003927	-0.83	0.404	-16.7785	6.756465
OBS	**-1.881221**	.7427958	-2.53	0.011	-3.337074	-.425368
$NetInt$	**-9.345881**	1.922006	-4.86	0.000	-13.11294	-5.578819
$Funding$	**-2.789421**	1.262326	-2.21	0.027	-5.263535	-.3153077
$\frac{1}{2}\ln(Assets)^2$	**-.0860225**	.0066432	-12.95	0.000	-.0990429	-.073002
t	.087296	.5258422	0.17	0.868	-.9433358	1.117928
$\frac{1}{2}t^2$	-.0135856	.160738	-0.08	0.933	-.3286262	.301455

To more clearly interpret the results in Table 19, it should be noted that explanatory variables defined above are correlated with profit inefficiency as specified in equation ((59)). Therefore, a negatively correlated variable with profit *inefficiency* can be interpreted as being positively correlated with profit *efficiency*. Estimates highlighted in bold typeface are those found significant at 1% or 5% critical levels.

Interestingly, results show that the ex post measure of credit risk ($CrdtRsk$) which is calculated as the ratio of loan loss provisions (LLPs) to total loans, seems to have the most significant and positive impact on European banks profit inefficiency (given its relatively large magnitude or coefficient). This can alternatively be viewed as ($CrdtRsk$) having a significant and negative impact on profit efficiency. The positive correlation of $CrdtRsk$ with profit inefficiency can always be expected, because the larger the level of loan losses (in terms of higher LLPs), the lower the profits and therefore the less profit efficient a given bank is. This is because LLPs are accounted for as expenses, and therefore the larger the value of LLPs (reflecting lower loans' quality), the lower the level of profits which translates into lower profit efficiency.

In fact, it is the notably large magnitude of the credit risk impact on profit efficiency that is worth contemplating. The large magnitude of this ex post credit risk measure indicates to a highly-sensitive risk-return tradeoff due to the considerable sensitivity of profit efficiency to the variation in credit risk. The risk-return tradeoff here suggests that for a slight increase in credit risk (in terms of increasing LLPs relative to loans), profit efficiency can experience a considerable and significant deterioration. Conversely, a slight decrease in credit risk (in terms of declining LLPs relative to loans) can translate into a considerable and significant improvement in profit efficiency.

This in turn suggests that maintaining the quality of the loans portfolio for European banks is vital to making profits more efficiently. This is despite that summary statistics are showing that the ratio of LLPs to total loans ($CrdtRsk$) has a maximum value of about 6.5% and a mean of under 1% (Table 3, page 199). In absolute terms, such a low ratio may indicate very low credit risk involved, however as the ratio of LLPs to total loans reflects the downside of credit risk, the low ($CrdtRsk$) ratio therefore suggests a favorable risk-return payoff given the low level of LLPs relative to the size of the loan portfolio. Looking into this differently, this low ($CrdtRsk$) ratio suggests that the size of total loans (denominator) is relatively too big relative to the level of loan losses (numerator). This indicates that a considerable credit risk was taken by European banks over the study period which seems to pay off given the very low level of defaults as represented by the corresponding low level of LLPs.

Accordingly, the relatively big magnitude of the impact of the ($CrdtRsk$) implies that profit efficiency is very sensitive to the quality of the loans portfolio, and the very low average value of ($CrdtRsk$) in absolute terms suggests a favorable risk-return payoff as the credit risk taken seem to have translated into very low level of loan losses.

On the other hand, trading risk ($TrdgRsk$), calculated as the ratio of net trading revenues (trading revenues – trading costs) to total revenues, is showing a significant and negative impact on profit inefficiency. The absolute value of the trading risk ratio is indicative of the level of exposure to trading regardless of whether it is positive (reflecting trading profits) or negative (reflecting trading losses), whereas the sign of the trading risk ratio reflects the risk-return pay-off in trading. The estimation result is logical as more net trading revenue suggests more profits and higher profit efficiency ultimately. In other words, taking on more trading risk, in the sense of achieving more net trading revenue relative to the bank's total income, results in higher profit efficiency for the sample of European banks involved. This also suggests that trading revenue is an important source of profitability given its significant impact on European banking profit efficiency.

Considering the positive impact of ($TrdgRsk$) on profit efficiency in conjunction with its negative impact on cost efficiency may seem contradicting, whereas it is not the case for two reasons. First, as will be seen later in this chapter, there is no significant correlation between profit and cost efficiencies, therefore drivers of cost and profit efficiencies can have different effects on both efficiencies. Second, the negative correlation of $TrdgRsk$ with cost efficiency suggests that, as banks take on more trading risk, increasing trading costs (and ultimately trading losses) seem to be incurred for a given level of output in trading, however considering this in line with its positive correlation with profit efficiency suggests that taking on more trading risk is also associated with increasing revenues from trading which translates into a positive impact on profit efficiency. The net impact for taking on more trading risk on technical efficiency, after all, seems to be in favor of greater profit efficiency. This is based on estimation results where $TrdgRsk$ is found with a larger and significant impact of on profit inefficiency (a coefficient of -12.2112) compared to its impact on cost inefficiency (a coefficient of 8.940631 as shown in Table 17).

Analyzing the impact of funding structure ($Funding$), which is calculated as the ratio of customer deposits to total funding, serves two purposes: it tests the impact of relying on customer deposits relative to other forms of funding on profit efficiency, and it also serves the purpose of isolating the funding-mix effect embodied in the liquidity risk ratio which will be discussed next. Estimation results for the variable ($Funding$) show a negative and significant correlation with profit inefficiency of (-2.789421), implying that: the greater the reliance on customer funding, hence the larger the ($Funding$) ratio, the better the ability of European banks to make profits more efficiently. This is because relying on a wider base of customer deposits as a main source of funding can boost profits as funding costs can be reduced because customer deposits is a relatively cheaper source of funding. This can ultimately result in enhancing profit efficiency.

Greater reliance on customer deposits to support the bank's assets is not only important to improve profitability as results here show, but this also has wider systematic implications. This is because banks with wide and diversified deposit bases can more effectively withstand pressures on liquidity in times of financial uncertainty. The recent example of Northern Rock is clear evidence on the vital role customer deposits have.

Liquidity risk ($LqdtyRsk$), calculated as the ratio of customer and short-term funding to liquid assets, is also showing a significant and negative impact on profit inefficiency. This finding suggests that taking more liquidity risk is associated with enhanced profit efficiency. This result can be interpreted from two perspectives as higher exposure to $LqdtyRsk$ can be the result of: (1) holding proportionately fewer liquid assets (the denominator effect), or (2) using more customer and short-term funding (the numerator effect). In this context, since liquid assets are the least revenue-generating assets on the bank's balance sheet, holding fewer liquid assets therefore suggests that more funds can be invested in more profitable assets which ultimately translate to enhanced profit efficiency.

The alternative explanation comes from looking into the numerator-effect, that is, taking more liquidity risk in terms of increasing reliance on customer and short-term funding which seems to be associated with enhanced profit efficiency. This result should only be interpreted in terms of increasing reliance on customer deposits as opposed to wholesale short-term funding. This is because such funding-mix effect is disentangled by two findings: (a) the positive correlation found between the funding ratio ($Funding$ = customer deposits / total funding) and profit efficiency which implies that customer funding is a cheap source of funding, and (b) the qualitative evidence in that money market funding is more expensive than customer funding (deposits) according to the ECB's financial stability report (ECB, 2004a). Consequently, with the funding-mix effect being disentangled, the conclusion that can therefore be derived given the estimation result is that: taking on more liquidity risk by increasing the reliance on customer deposits and reducing liquid asset holdings improves profit efficiency. It should be noted however that the impact of liquidity risk is found with the least magnitude on profit inefficiency (a coefficient of -.0212271) compared to other significant factors.

Further, estimation shows a negative and significant influence of OBS (the credit equivalent of off-balance sheet items / total assets) on profit inefficiency (a coefficient of -1.881221). Such finding suggests that expansion in OBS activities enhances profit efficiency for European banking. Therefore, the greater the involvement in OBS activities, the greater the ability of European banks to make profits more efficiently. In comparison with recent past research, this result is in contrast to the finding of Pasiouras (2008) using Greek banking data over the period of 2008 – 2018, where the author found that the inclusion of OBS items as an output does not have any impact on the efficiency scores.

Notwithstanding the significant impact of the OBS activities on profit efficiency, income generated by traditional lending business remains a major factor driving profit efficiency for European banks. Findings indicate a very significant and negative impact at the 1% critical level of $NetInt$, calculated as net interest income to total income, on profit inefficiency (-9.345881). That is to say, the greater the proportion of net interest income to total income, the better the ability of European banks to make profits more efficiently. However, by comparing the influence of non-interest versus interest-related incomes on profit efficiency as results suggest, it is interesting to note that the impact of interest income ($NetInt$) is second in importance to that of net trading income ($TrdgRsk$) given that the magnitude (estimated coefficient) of the latter is greater than that of the former, where both are statistically significant at the 1% critical level.

It is interesting to reflect on the estimation outcome related to the ex ante credit risk measure, that is, credit risk appetite CRA. Results show that CRA (that is calculated as the ratio of risk weighted assets RWA to total assets) has no significant impact on European banks profit efficiency. That is, the size of risk weighted assets (RWA) reflecting the level of credit risk exposure relative to bank size (total assets) is showing no significant influence on profit efficiency. Moreover, CRA was also found to be insignificantly correlated with European banks cost efficiency. Such a result may indicate that CRA is an insensitive credit risk measure in terms of reflecting the outcome of the risk-return payoff for European banks over 2008 – 2018. This is possibly because CRA is based on Basel I risk weighting criteria that has long been criticized for having the problem of one-size-fits-all. Had the percentage of RWA to total assets been representative of the bank’s ex ante credit risk exposure, results could have shown either a significantly positive or negative correlation with profit or cost efficiencies reflecting the state of the risk-return payoff.

Both cost and profit efficiency estimation results related to (CRA) represent good evidence of the limitedness of Basel I and the need for more comprehensive and risk-sensitive regulatory framework. The Economist (Nov. 22nd 2007, p 31) confirms this view by arguing that regulators may tighten capital requirements “as confidence in the tier-one capital ratio favoured by European regulators seems to have evaporated…investors seem to pay closer attention to more cautious capital measures which does not allow for any risk-weighted adjustment to assets”. The Committee of European Banking Supervisors CEBS, which is responsible for setting the regulatory framework for Europe's evolving single market in financial services, have realized the increasing complexity of the banking business and the need for a more developed framework. This is clearly recognized by the head of the CEBS as it is stressed that "banking used to be a very easy business 20 or 30 years ago. It was a plain vanilla business. You only have to look at Basel II and the risk management techniques to realise that the complexity of the business has increased and the complexity of the supervisory approaches has increased parallel to that" (The Financial Times, Apr 20th 2004).

Estimation result for the total assets variable ($\frac{1}{2}\ln Assets^2$) shows a significant and negative correlation with profit inefficiency at 1% level (a coefficient of -.0860225). Such finding is indicative of a significantly positive size-profit efficiency relationship which suggests that the larger the bank the greater its ability to make profits more efficiently. This result is possibly supported by the significantly positive correlation earlier found between the size of off-balance sheet assets and European banks profit efficiency. The expansion in off-balance sheets activities (OBS) therefore may be one reason causing this positive size-profit efficiency relationship considering that, on average, around 1/3 (28%) of European banks' assets over 2010 -2018 were held as OBS items. To further investigate this size-profit efficiency relationship, a correlation (covariance) test was conducted between observation-specific bank sizes (assets) and profit efficiency scores. Interestingly, the test reveals quite a moderate relationship (a correlation coefficient of 0.26) but with considerable statistical significance at the 1% critical level (p-value = 0.0004). To shed some more light on this issue, the following subsection classifies sample banks into 5 asset groups and provides group-specific profit and cost efficiency scores.

5.5.6 Profit Efficiencies by Asset Group

The aim of this exercise is to further explore the significant correlation found between bank size and profit efficiency. Sample banks are classified into 5 asset groups and efficiency scores were re-produced as group-specific averages as shown in the table below:

Table 20: Profit Efficiencies by Asset Group

Asset Group	**Profit Efficiency**		
	Freq.	Mean	Std. Dev.
AG1: assets < €15bn	47	.8104252	.2447349
AG2: €15bn ≥ assets < €50	141	.9503112	.0614937
AG3: €50 ≥ assets < €150bn	83	.9639498	.0406772
AG4: €150bn ≥ assets < €400bn	241	.9819510	.0187410
AG5: assets ≥ €400bn	29	.9806375	.0322163
Sample Mean	*541*	*.9368076*	*.1306509*

The positive correlation between bank size and profit efficiency found in the estimation is further confirmed by the results shown in the table above. It is clear that profit efficiency is improving as banks get larger, with AG4 and AG5 banks demonstrating the highest levels of profit efficiency of about 98%. Notably, AG4 banks are shown to be the most profit efficient with the lowest standard deviation in the sample of just under 2%. The conclusion that can be drawn here is that larger banks are more profit efficient than smaller banks, where banks with total assets between €150bn and €400bn are shown to be the most profit efficient.

Previous European and US studies found mixed evidence on the size-profit efficiency relationship. Maudos et al (2002) using banking data from 10 EU countries over 1993 – 1996 find a non-linear relationship between bank size and profit efficiency, as only medium size banks (up to $10bn) were found with a negative and significant correlation with profit efficiency. On the other hand, Vennet (2002) uses a sample from 17 European countries over 1995 – 1996 and finds little evidence on size-profit efficiency relationship. Likewise, Kasman and Yildirim (2006) use SFA in a single-stage estimation approach to estimate profit function (truncated at the 2nd order terms) for commercial banks in 8 new member states of the EU in 2004 and find no significant relationship between size and profit efficiency.

Moreover, Berger et al (1993a) using US data over 1984 – 1989 find strong and positive correlation between profit efficiency and size. Berger and Mester (1997) use US data over 1990 – 1995 and find a negative correlation between size and profit efficiency. Akhigbe and McNulty (2005) use SFA and the two-stage approach to estimate year-wise alternative profit function over the period of 1995 – 2001 and also find a positive size-efficiency relationship. Further, Fitzpatrick and McQuinn (2008) estimate a Translog 'alternative' profit function for large commercial banks in Canada, the UK, Ireland and Australia over 1996 – 2002 and use SFA and the single-stage estimation approach (specifying Battese and Coelli 1995 inefficiency model) find a positive size-profit efficiency relationship.

Estimation results for profit efficiency determinants of European banks have revealed the following conclusions. For European banks to make profits more efficiently, they need to – in terms of priority (i.e. impact significance): (1) maintain high quality loan portfolio (2) better manage their trading risks (3) reduce liquid asset holdings and rely more on customer deposits, (4) focus on non-interest income and expand their OBS activities, and finally (5) expand in size, as results suggest that larger banks with total assets between €150bn and €400bn, can significantly make profits more efficiently as they are found with the highest profit efficiency in the sample.

5.5.6.1 Correlation Between Efficiencies

This study extends the analysis of bank-specific cost and profit efficiency estimates by investigating any potential correlation between the two efficiencies. To this end, Spearman test for ranks correlation is applied. The null hypothesis here states that the two efficiencies are independent. Empirical results show a correlation coefficient of (0.0287) that is statistically insignificant at 5% level ($P > |t| = 0.6969$). Thus, the null of no correlation between profit and cost efficiencies cannot be rejected. The conclusion therefore is that there is no significant correlation between European banks' profit and cost efficiencies as far as this sample data is concerned. This finding is consistent with that of Berger and Mester (1997) using US data.

Now, in order to provide some further insights into profit and cost efficiency results, the progression of profit and cost efficiencies over time is analysed. The aim is to see if more conclusions can be drawn from this analysis results which may in turn imply important policy implications.

CHAPTER 7: CONCLUSION

7.1 Summary and Conclusion

This research has yielded findings that can be recognized as a genuine contribution to knowledge in the area of banking efficiency analysis. To elaborate, this research results are presented from three main perspectives: functional form modification outcome, frontier estimation results and inefficiency explanatory variables estimation results.

First, the systematic modification of the functional form has proved to be pivotal to conduct prior to the estimation of the efficiency frontier. Results clearly show that, not only it is statistically vital to incorporate the fourier flexible terms, but also its equally important to include risk factors (on a country level) in the fabric of both the cost and profit functional forms (as shown Table 12 and **Error! Reference source not found.**). Accordingly, the first well-founded finding of this research can be articulated as follows: *ignoring this risk-modified fourier flexible functional forms modification is most likely to produce misleading estimation results*.

Second, cost frontier estimation results have shown that banks operating in EU-economies are found to incur higher costs than banks operating in non-EU-economies, thus it is less cost efficient to operate in an EU economy over the study period. Moreover, results have also shown that banks' costs increase with country-specific credit, trading, and liquidity risks, indicating that total costs increase with (1) loan losses, (2) trading costs (that are translating into trading losses), and (3) money market (or short term) funding. This implies that *credit, trading and liquidity risks are not well managed on the Macro-economic level over the study period of 2008 – 2018 causing banks to incur higher cost to operate*.

Third, the estimation results regarding the determinants of technical cost inefficiency indicate that as banks take more trading risk, i.e. as net trading revenue increases relative to total revenues, European banks become less cost efficient. Such result is in line with the negative impact of the country-level trading risk on banking costs, suggesting that as banks take more trading risk, they incur more costs and become more cost inefficient accordingly, *indicating that trading risks should be re-visited on a* *bank and policy makers' levels*. Results also confirm that European banks can operate more cost efficiently when banks are better capitalized. On the other hand, there is no significant evidence supporting the correlation of cost efficiency with the level interest margin (relative to total revenues) nor that with bank size either. This suggest that policy makers may consider the fact that *better capitalizing banks would assist them in becoming more profit efficient, regardless of their size*.

Fourth, profit frontier estimation results suggest that profits tend to increase if banks operate in a non-euro-economy, whereas banks' profits experience an upward shift for if they are operating in the EU. Further, results show that profits are positively associated with inflation rate and negatively correlated with short-term interest rates, a result that is economically justified. Risk-wise, results clearly show a negative correlation between country-level liquidity risk and banks' profitability, such that the lower the liquidity risk (in the sense of using less money market funding) the higher the profitability, suggesting that less reliance on money market funding leads to greater profits.

Fifth, the estimation results regarding the determinants of technical profit inefficiency indicate that, for European banks to make profits more efficiently, they can support the following six recommendations. First of all, maintain high quality loan portfolios given the considerable sensitivity of profit efficiency to credit risk, second, better manage trading risk to enhance trading risk-return payoff, since trading risk measure is found to significantly and negatively correlate with profit inefficiency. Third, banks need to rely more on customer deposits relative to total funding, as the funding ratio is showing a significantly negative correlation with profit inefficiency.

In addition, banks need to reduce liquid asset holdings as liquidity risk is negatively correlated with profit efficiency, and/or rely more on customer funding as shown in the previous point, to significantly increase profit efficiency. Moreover, banks should consider expanding their OBS activities, given the negative and significant correlation of the OBS ratio (credit-equivalent of OBS items/total assets) with profit inefficiency. Finally, focus on interest income and therefore expand their lending activities, given the significant impact of net interest income as a proportion of total revenue on profit efficiency. For policy implications, *banks can be more profit efficient as they better manage their loan portfolios in terms of diversity and quality, reduce their trading activities and focus more on their core business of attracting customer deposits as the main source of funding and making loans. Banks are also advised to hold fewer liquid assets, yet expand their OBS activities*.

Sixth, concerning the Credit Risk Appetite (CRA), which is used as a proxy for bank-specific credit risk, results show that it has no significant impact on European banks profit efficiency. Equally, CRA is also found to be insignificantly correlated with European banks cost efficiency. These two results together seem to indicate that is an insensitive credit risk measure as it is a poor indicator of the level of credit risk exposure, hence it seems to fail to provide an indication of the state of the risk-return payoff for European banks over 2008 – 2018. This is possibly because seems to be somehow detached from real credit risk exposure taken on by banks. The research implication of this result is that *further studies should consider more reflective measure of credit risk than the CRA*.

Finally, results show that there are slightly more profit inefficiencies than cost inefficiencies in European banking. This is consistent with the finding of the majority of past European, US, and International research. Average profit efficiency is around 93.68% (or an average profit inefficiency of about 6.32%), whereas average cost efficiency is found at about 1.050 (or an average cost inefficiency of about 5%). Lastly, the analysis shows that there is no significant correlation between bank-specific cost and profit efficiencies. For policy implications, *the priority for European banks is to focus on the factors and procedures that minimize their profit inefficiencies, and then take actions that minimize their cost inefficiencies*.

7.2 Implications for Policy Makers and Further Research

Given the above-mentioned research findings, a number of policy and research implications can be derived accordingly. European banking policy makers may consider a set of financial regulations in the aim of assisting European banks to encourage banks to have more robust capitalization and better manage credit, trading and liquidity risks on the Macro-economic level, specifically, trading risks, which should be minimized, regardless of their size. Banks should improve their loan portfolio management, hold less liquid assets, widen their customer deposit base, expand their off-the-balance sheet activities, yet, downsize their trading activities. Lastly, the priority for European banks is to minimize their profit inefficiencies, and then take actions that minimize their cost inefficiencies.

Finally, to pave the way for more robust bank-efficiency studies in the future, risk-modified fourier flexible functional forms must be considered prior to estimating efficiency scores, as ignoring which is most likely to produce misleading estimation results. Further studies should also consider more representative measure of credit risk than the CRA.

REFERENCES

Akhigbe, A., and McNulty, J. E., 2003. The profit efficiency of small US commercial banks. Journal of Banking and Finance 27, 307 – 325.

Akhigbe, A., and McNulty, J. E., 2005. Profit Efficiency Sources and Differences among Small and Large U.S. Commercial Banks. Journal of Economics and Finance, 29, 3, 289 – 299.

Aigner, D., Lovell, C., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production models. Journal of Econometrics 6, 21 – 37.

Allen, L., Rai, A., 1996. Operational efficiency in banking: An international comparison. Journal of Banking and Finance, 20, 655 – 672.

Allen, L., Boudoukh, J. and Saunders A., 2004. Understanding Market, Credit, and Operational risk: The value at risk approach. Blackwell Publishing.

Altunbas, Y., and Molyneux, P., 1994. Sensitivity of stochastic frontier estimation to distributional assumptions: the case of the German banks. Institute of European Finance paper (preliminary version, unpublished).

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

Altunbas, Y. and Chakravarty S. P., 2001. Frontier cost functions and bank efficiency. Economics Letters 72, 233 – 240.

Altunbas Y., Gardener, E., Molyneux, P., and Moore, P., 2001. Efficiency in European Banking. European Economic Review, 45, 1931 – 1955.

Altunbas, Y., Carbo, S., Gardener E.P.M. and Molyneux, P., 2003. Ownership and performance in European and US banking – A comparison of commercial, cooperative and savings Banks, mimeo (University of Wales, Bangor, 2003).

Altunbas, Y., Carbo, S., Gardener, E.P.M and Molyneux, P. 2007. Examining the Relationship between Capital, Risk and Efficiency in European Banking. European Financial Management 13, 1, 49 – 70.

Amel, D., et al, 2004. Consolidation and efficiency in the financial sector: A review of the international evidence. Journal of Banking and Finance, 28, 2493 – 2519.

Ariff, M, and Can, L., 2008. Cost and profit efficiency of Chinese banks: A nonparametric analysis. China Economic Review, 19, 260 – 273.

Athanassopoulus, A. D., 1995. Multivariate and frontier analysis for assessing the market and cost of large-scale bank branch networks. Warwick Business School Research Bureau, 168.

Ausina, E. T., 2002. Bank Cost Efficiency and Output Specification, Journal of Productivity Analysis, 18, 199 – 222.

Avila, D. R., 2007. Finance and growth in the EU: New evidence from the harmonization of the banking industry. Journal of Banking and Finance 31, pp 1937 – 1954.

Bank of International Settlements BIS, 1999. Capital Requirements and Bank Behavior: The Impact of the Basle Accord. Basel Committee on Banking Supervision, No.1 April 1999.

Bank for International Settlements BIS, 2003. Overview of the New Basel Capital Accord: Consultative Document. Basel Committee on Banking Supervision, April 2003.

Bank of England, 2002. Bank Capital: Basel II Developments, Financial Stability Review, December 1992, 103 -109.

Bank of England, Dec 2003. The financial stability conjuncture and outlook, Financial Stability Review, December, 66-79.

Bank of England, 2005. Financial stability: Managing liquidity risk in global system, The 14th City of London Central Banking & Regulatory Conference, November 28th.

Banker, R. D. 1993. Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation. Management science, 39 (10), 1265-1273.

Barclays PLC 2006. Annual Report.

Basel Capital Accord, 1988. International Convergence of Capital Measurements and Capital Standards. Bank of International Settlements, Basle Committee July 1988.

Battese, G.E. and Coelli, T. J., 1992. Frontier Production Functions, Technical Efficiency and Panel Data: With application to Paddy Farmers in India. Journal of Productivity Analysis, 3, 153-169.

Battese, G.E. and Coelli, T. J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data, Empirical Economics, 20, 325-332.

Battese, G. E., et al 2004. A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies. Journal of Productivity Analysis 21, 91 -103.

Barros, C. P., et al, 2007. Analyzing the determinants of performance of best and worst European banks: A mixed logit approach. Journal of Banking and Finance 31, 2189 – 2203.

Bauer, P. W., Berger A. N., and Humphrey D. B. 1993. Efficiency and Productivity Growth in U.S. Banking. The Measurement of Productive Efficiency: Techniques and Applications, (Oxford: Oxford University Press), 386 – 413.

Bauer, Paul W., and Ferrier, G. D., 1996. Efficiency Measurement Issues for Payments Processing. Journal of Money, Credit and Banking 28, 1004 – 39.

Bauer, P.W., Berger, A.N., Ferrier, G.D., Humphrey, D.B., 1998. Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. J. Econ. Business 50, 85–114.

Baum, C. F., 2006. An introduction to modern econometrics using Stata, StataCorp LP, Stata Press.

Beccalli, E., 2007. Does IT investment improve bank performance? Evidence from Europe. Journal of Banking and Finance 31, 2205 – 2230.

Bell, F. W., and Murphy, N. B. 1968. Costs in Commercial Banking: A Quantitative Analysis of Bank Behavior and its Relation to Bank Regulation. Research Report No. 41, Federal Reserve Bank of Boston.

Berg, S. A., et al, 1991, Technical efficiency of Norwegian banks: A nonparametric approach to efficiency measurement, Journal of Productivity Analysis 2, 127 – 142.

Berger, A. N. and Hannan, T.H., 1989, The price-concentration relationship in banking, Review of Economics and Statistics 71, 291 – 299.

Berger, A. N., and Humphrey, D., 1992. Measurement and Efficiency Issues in Commercial Banking. In Z. Griliches (ed.), Output Measurement in the Service Sector, Chicago: The University of Chicago Press, pp 245-300.

Berger, A. N., et al, 1993a. Bank efficiency derived from the profit function. Journal of Banking and Finance, 17, 317 – 347.

Berger, A. N. et al, 1993b. The efficiency of financial institutions: a review and preview of research past, present and future. Journal of Banking and Finance, Vol. 17, 221 – 249.

Berger, A. N., 1995. The relationship between capital and earnings in banking. Journal of Money, Credit and Banking, Vol. 27, No. 2, pp 432-456.

Berger, A. N., and De Young, R., 1997. Problem loans and cost efficiency in commercial banks. Journal of Banking and Finance, 21, 849 – 70.

Berger, A. N. and Humphrey, D., 1997. Efficiency of Financial Institutions: International Survey and Directions for Future Research. European Journal of Operational Research, 16 April 1997, vol. 98, issue 2, 175 – 212.

Berger, A. N., et al, 1997. The Efficiency of Bank Branches. Journal of Monetary Economics, 40, 141 – 162.

Berger, A. N. and Mester, L. J. 1997. Inside the black box: What explains differences in the efficiencies of financial institutions? Journal of Banking and Finance 21, 895 – 947.

Berger, A. N., 2000. The Integration of Financial Services Industry: Where are the Efficiencies? North American Actuarial Journal, Vol. 4, No. 3, 1 – 28.

Berger, A. N., et al, 2000. Globalization of financial institutions: Evidence from cross-border banking performance. Brookings-Wharton papers on financial services, 3rd ed. Washington, DC: Brookings Institution.

Berger, A. N., De Young, R., and Udell G. F., 2001. Efficiency barriers to the consolidation of the European financial services industry. European Financial Management. Vol. 7, No. 1, 117 – 130.

Berger, A. N. and Mester L. J. 2003. Explaining the dramatic Changes in Performance of US Banks: Technological change, deregulation, and dynamic changes in competition. Journal of financial intermediation 12, 57 – 95.

Berger, A. N., et al, 2007. Competition from Large, Multimarket Firms and the Performance of Small, Single-Market Firms: Evidence from the Banking Industry. Journal of Money, Credit and Banking 39, 331 – 368.

Berndt, E. R., and Khaled M. S., 1979. Parametric Productivity Measurement and Choice Among Flexible Functional Forms. Journal of Political Economy, 87, 1220 – 1245.

Besanko, D. and Braetigam, R. R., 2005. Microeconomics. 2nd edition, John Wiley and Sons.

Bos J. W. B., and Kolari, J. W., 2005. Large Bank Efficiency in Europe and the United States: Are There Economic Motivations for Geographic Expansion in Financial Services? Journal of Business, 78, 4, 1555 – 1592.

Bos, J. W. B., and Schmiedel, H., 2007. Is there a single frontier in a single European banking market? Journal of Banking and Finance 31, 2081 – 2102.

Brazdik, F., 2005. Oriented Stochastic Data Envelopment Analysis Models: Ranking Comparison to Stochastic Frontier Approach. Charles University, Center for Economic Research and Graduate Education, Academy of Sciences of the Czech Republic, August 2005, Working Paper.

Brealey, R., A., and Myers, S., C. 1996. Principles of corporate finance. 7th edition, McGraw Hill.

Bronson, R., 1989. Matrix Operations. Schaum's Outline Series in Mathematics, McGraw-Hill Book Company.

Brooks, C., 2002. Introductory Econometrics for Finance. Cambridge University Press.

Casu, B. and Molyneux, P., 2003. A comparative study of efficiency in European banking, Applied Economics 35, 1865 – 5416.

Casu, B. and Girardone, C., 2004. Large Banks' Efficiency in the Single European Market. The Service Industries Journal 24, 6, November 2004, 129 – 142.

Casu, B. and Girardone, C., and Molyneux, P., 2004. Productivity change in European banking: A comparison of parametric and non-parametric approaches, Journal of Banking and Finance 28, 2521 – 2540.

Caves, D. W., Christensen L. R., and Tretheway, M. W., 1980. Flexible Cost Functions for Multiproduct Firms, The Review of Economics and Statistics, 62, 477 – 481.

Choudhry, M., 2007. Bank Asset and Liability Management: Strategy, Trading, Analysis. John Wiley & Sons (Asia) Ltd.

Christensen, L. R., Jorgensen, D. W., and Lau, L. J., 1973. Transcendental Logarithmic Production Frontiers, Review of Economics and Statistics, Vol. 55, issue 1, pp 28-45

CIMA Official Terminology 2005. Elsevier & CIMA Publishing, London.

Cark, J. and Siems, T., 2002. X-efficiency in banking: looking beyond the balance sheet. Journal of Money, Credit, and Banking, 34, 987 – 1013.

Claudia, G., Molyneux, P., and Gardner, E. P. M., 2004. Analyzing the determinants of bank efficiency: the case of Italian banks, Applied Economics, 36, 215 – 227.

Coelli, T., 1996. A Guide to FRONTIER Version 4.1: A computer program for Stochastic Frontier Production and Cost Function Estimation. Centre for Efficiency and Productivity Analysis, Working Paper 96/07.

Coelli, T., Prasada, R., Battese, G. E., 1998. An introduction to efficiency and productivity analysis, Kluwer Academic Publishers.

Charnes, A., Cooper, W. W., and E. Rhodes, 1978. Measuring the Efficiency of Decision-Making Units. European Journal of Operational Research, Vol. 2, 429 – 44.

Danthine, J. P., Giavazzi, F., Vives, X., and Thaaden, E. V., 1999. the future of European Banking, Center for Economic Policy Research, Monitor8ng European Integration Paper No. 9, London: CEPF, 1999.

De Prins, D., Simar, L., & Tulkens, H., 1984. Measuring labour efficiency in post offices. In M. Marchand, P. Pestieau & H. Tulkens (Eds.), The performance of public enterprises: Concepts and measurement, 243 – 267. Amsterdam: North Holland.

De Young, R., 1997. Measuring Bank Cost Efficiency: Don't Count on Accounting Ratios, Financial Practice and Education.

Dermine, J., 2003. Banking in Europe: Past, present and future. Proceedings of the 2nd ECB Central Banking Conference on the Transformation of the European Financial System. ECB Frankfurt.

Dermine, J., 2006. European banking integration: Don't put the cart before the horse. Financial Institutions, Markets and Money, 15, 57 – 106.

Deutsche Bank, 2006. Annual Review.

Diewert, W. E., and Kopp, R. J., 1982. The decomposition of frontier cost function deviations into measures of technical and allocative efficiency. Journal of Econometrics 19, 319 – 331.

Dowd, K., 2002. An introduction to market risk measurement. Wiley Finance.

Dowd, K., 2007a. Letters to the Editor: Better to have thrown Northern Rock to the wolves, The Financial Times September 19th 2007.

Dowd, K., 2007b. Letters to the Editor: Northern Rock managers like captain of Titanic, The Financial Times October 19th 2007.

Drake, L., and B. Howcroft, 1993. Relative efficiency in the branch network of a UK bank: an empirical study. International Journal of Management Science 22, 1, 83 – 90.

Drake, L., and B. Howcroft, 1995. Measuring the Relative Efficiency of the Selling Function: An Application of Data Envelopment Analysis to UK Bank Branches, working paper, Loughborough University, U.K. (April 1995).

Drake, L., and Hall, M. J. B., 2003. Efficiency in Japanese banking: An empirical analysis. Journal of Banking and Finance 27, 891 – 917.

Drake, L., Hall, M. J. B., and Simper, R., 2006. The impact of macroeconomic and regulatory factors on bank efficiency: A non-parametric analysis of Hong Kong's banking system. Journal of Banking and Finance 30, 1443 – 1466.

Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models. Stata Journal 3, 168 – 177.

Eastwood, B. J., and Gallant, A. R., 1991. Adaptive Rules for Seminonparametric Estimators That Achieve Asymptotic Normality. Econometric Theory, 1, 307-40.

English, M., S. Grosskopf, K. Hayes and S. Yaisawarng, 1993. Output allocative and technical efficiency of the financial services sector, Journal of Banking and Finance 17, 349 – 366.

European Economy, 1988. Creation of a European financial area. Commission of the European Communities 36, EU, Brussels.

Fare, R., et al, 2004. The effect of risk-based capital requirements on profit efficiency in banking. Applied Economics, 36, 1731 – 1743.

Fenn, P., Vencappa, D, Diacon, S., Klumpes, P., and O'Brien, C., 2008. Market structure and the efficiency of European insurance companies: A stochastic frontier analysis. Journal of Banking and Finance, 32, 86 – 100.

Ferrier, G. D., Lovell, C. A. K., 1990. Measuring cost efficiency in banking: Econometric and linear programming evidence. Journal of Econometrics 46, 229 – 245.

Fiorentino, E., Karmann, A., and Koetter, M., 2006. The cost efficiency of German banks: a comparison of SFA and DEA. Discussion Paper Series 2: Banking and Financial Studies, No 10/2006. Deutsche Bundes Bank.

Fitzpartick, T., and McQuinn, K., 2008. Measuring bank profit efficiency. Applied Financial Economics, 18, 1 – 8.

Flannery, M.J., 1984. The social costs of unit banking restrictions. Journal of Monetary Economics 13, 237 – 249.

Frerier, G D., Lowell, C.A.K 1990. Measuring cost efficiency in banking: Econometric and linear programming evidence. Journal of Econometrics 46, 299 – 245.

Fried, H. O., Lovell C. A. K. and Schmidt, S. S., 1993. The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press.

Frost, S. M., 2007. The Bank Analyst's Handbook: Money, risk and conjuring tricks. John Wiley and Sons, Ltd.

Gallant, A. R., 1981. On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form. Journal of Econometrics, 15, pp 211- 45.

Gallant, A. R., 1982. Unbiased Determination of Production Technologies. Journal of Econometrics 20, 285 – 323.

Gilbert, R. A., 1984. Bank Market Structure and Competition: A Survey, Journal of Money, Credit, and Banking, 16, 617 – 645.

Girardone, C., Molyneux, P., and Gardner, E. P. M., 2004. Analyzing the determinants of bank efficiency: the case of Italian banks. Applied Economics 36, 215 – 227.

Goddard, J., Molyneux, P., Wilson, J., 2004. Dynamics of Growth and Profitability in Banking. Journal of Money, Credit, and Banking, 36, 6, 1069 – 1090.

Goddard, J., Molyneux, P., Wilson, J., Tavakoli, M. 2007. European Banking: An Overview. Journal of Banking and Finance, Vol. 31, 1911 – 1935.

Gould, W., and Sribney, W., 1999. Maximum Likelihood Estimation with Stata. Stata Corporation StataCorp.

Greene, W. H., 2000. Econometric Analysis. 4th edition, Prentice Hall.

Gropp, R., et al, 2002. Equity and bond market signals as leading indicators of bank fragility. European Central Bank working paper no. 150.

Gropp, R., Moermann, G.A., 2004. Measurement of contagion in banks equity prices. Journal of International Money and Finance, 23, 405 – 419.

Grosskopf, S., 1996. Statistical Inference and Nonparametric Efficiency: A Selective Survey, The Journal of Productivity Analysis, 7, 161 – 176.

Gup, B. E., and Kolari, J. W., 2005. Commercial banking: The management of risk. 3rd edition, Wiley.

Gutierrez, R. G., Carter, S. L., and Drukker, D. M, 2001. On boundary-value likelihood ratio tests. Stata Technical Bulletin, 60, 15 – 18.

Hair et al, 1998. Multivariate data analysis. 5th edition, Prentice Hall.

Hamilton, L. C., 2003. Statistics with STATA: Updated for Version 8. Duxbury Press; 5th edition, Thomson.

Hancock, D. 1991. A Theory of Production for the Financial Firm. Kluwer Academic Publishers.

Hannan, T.H., 1991. Bank commercial loan markets and the role of market structure: Evidence from the survey of commercial lending, Journal of Banking and Finance 15, 133 – 149.

Harris, J. W. and Stocker, H. 1998. Handbook of Mathematics and Computational Science. New York: Springer-Verlag.

Heffernan, S., 2002. Modern Banking in Theory and Practice, John Wiley and Sons.

Hellmann, T.F., Murdock, K.C., and Stiglitz, J. E. Liberalization, 2000. Moral hazard in banking and prudential regulation: are capital requirements enough? American Economic Review, 90, 147-165.

Hempel, G. H., and Simonson, D., G. 1999. Bank Management: Texts and Cases, John Wiley and Sons.

HSBC Holding plc, 2006. Annual Report and Accounts.

Hughes, J. P., Mester L. J. 1993. A quality and risk-adjusted cost function for banks: evidence on the “too-big-to-fail” doctrine. Journal of Productivity Analysis 4, 293 – 315.

Hughes, J.P., Lang, W., Mester, L. J., and Moon, C.G. 1995. Recovering banking technologies when managers are not risk-neutral. Conference on Bank Structure and Competition. Federal Reserve Bank of Chicago May 1995, 349 – 368.

Hughes, J.P., Lang, W., Mester, L.J., and Moon, C.G. 1996a. Efficient banking under interstate branching. Journal of Money, Credit, and Banking 28, 1045 – 1071.

Hughes, J.P., Lang, W., Mester, L.J., and Moon, C.G. 1996b. Safety in numbers? Geographic diversification and bank insolvency risk. Working paper 96 – 14, Federal Reserves Bank of Philadelphia.

Humphrey, D. B. 1987. Cost dispersion and the measurement of economies in Banking. Economic Review, May/June, pp 24 – 38.

Humphrey, D.B. and Pulley, L.B., 1997. Banks' responses to deregulation: Profits, technology, and efficiency. Journal of Money, Credit, and Banking 29, 73 – 93.

Hunter, W.C. and S.G. Timme, 1986. Technical change, organizational form, and the structure of bank productivity. Journal of Money, Credit and Banking 18, 152 – 166.

Hunter, W.C. and S.G. Timme, 1991. Technological change in large U.S. commercial banking. Journal of Business, 64 (3), 339 – 362.

Hunter, W. C., Stephen G. T., and Won K. Y., 1990. An Examination of Cost Subadditivity and Multiproduct Production in Large U.S. Banks. Journal of Money, Credit, and Banking 22, 504 – 25.

Iannotta, G. et al. 2007. Ownership structure, risk and performance in the European banking industry. Journal of Banking and Finance 31, 2127 – 2149.

Isik, I, and Hassan, M. K. 2003. Efficiency, ownership and market structure, corporate control and governance in the Turkish banking industry. Journal of Business, Finance and Accounting 30, 1363 – 1421.

Iyer, K. G., Rambaldi, A. N. and Tang K. K., 2005. Measuring Efficiency externalities from Trade and Alternative Forms of Foreign Investment. Working Paper Series No. 04/2005, Centre for Efficiency and Productivity Analysis CEPA. School of Economics, University of Queensland, Australia.

Jacques, I., 2006. Mathematics for Economics and Business, 5th edition, Pearson Education Limited.

John G. R. and John G. 2008. Advanced Microeconomics. Cambridge University Press.

Jorgenson, D. W., 1986. Econometric Methods for Modeling Producer Behavior. Handbook of Econometrics, 3, 1841 – 1915.

Jorgenson, D. W., and Fraumeni, B. M., 1981. Relative Prices and Technical Change in Modelling and Measuring Natural Resources Substitution. Cambridge, Massachusetts, MIT Press.

JP Morgan Chase & Co. 2005. Annual Report and Accounts.

Kaparakis, E.I., Millar, S.M., Noulas, A.G., 1994. Short-run cost inefficiencies of commercial banks. Journal of Money, Credit, and Banking 26, 875 – 893.

Kasman, A., and Yildirim, C., 2006. Cost and profit efficiencies in transition banking: the case of new EU members. Applied Economics, 38, 1079 – 1090. Kennedy, P. 2003. A Guide to Econometrics. 5th edition, The MIT Press: Cambridge, Massachusetts.

Kumbhakar, S. C., Ghosh S., and McGukin J. T., 1991. A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. Journal of Business and Economics Statistics, 9, 279 – 286.

Kumbhakar, S. C. and Lovell C. A. K. 2000. Stochastic Frontier Analysis. Cambridge University Press.

Kwan, S., Eisebeis, R. A., 1997. bank risk, capitalization, and operating efficiency. Journal of Financial Services Research. 12, 117 – 131.

Le Compte, R. L. B., and Smith, S. D., 1990. Changes in the Cost of Intermediation: The Case of Savings and Loans, Journal of Finance, 45, 1337 – 1346.

Leibenstein, H., 1966, Allocative efficiency vs. ‘X-efficiency’, American Economic Review 56, 392 – 415.

Lensink, R., and Hermes, N., 2004. The short-term effects of foreign bank entry on domestic bank behavior: Does economic development matter. Journal of Banking and Finance 28, 553 – 568.

Li, S. X., 1998. Stochastic models and variable returns to scales in data envelopment analysis. European Journal of Operational Research, 104, 532 – 548.

Lindquist, K., 2004. Banks’ buffer capital: how important is risk. Journal of International Money and Finance 23, 493 – 513.

Lumby, S., and Jones, C., 2003. Corporate Finance: Theory and Practice, 7th edition, Thomson.

MaClaney, E, 2006. Business Finance: Theory and Practice, 7th ed. Pearson Education.

MacDonald, S. S., Koch T. W., 2006. Management of Banking. Thomson South-Western 6th ed.

Mankiw, G. N. and Taylor M. P., 2006. Microeconomics. London, Thomson.

Martin, A. D., and Mauer, L. J., 2003. Exchange rate exposures of US banks: A cash flow-based methodology. Journal of Banking and Finance 27, 851 – 865.

Maudos, J., Pastor, J. M., Perez, F., and Quesada, J., 2002. Cost and profit efficiency in European banks. Journal of International Financial Markets, Institutions and Money 12, 33-58.

Mckillop, D. G., et al, 1996. The composite cost function and efficiency in giant Japanese banks. Journal of Banking and Finance, 20, 1651 – 1671.

Meeusen, W. Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18, 435 – 444.

Mester L. J., 1987a. A multiproduct cost study of saving and loans. Journal of Finance, 42, 423 – 445.

Mester L. J., 1991. Agency costs among savings and loans. Journal of Financial Intermediation 3, 257 – 278.

Mester L. J., 1992. Traditional and non-traditional banking: an information theoretic approach. Journal of Banking and Finance 16, 545 – 566.

Mester L. J., 1993, Efficiency in the savings and loan industry. Journal of Banking and Finance 17, 267 – 286.

Mester, L., 1994a. How efficient are third district banks, Business Review, January, 3 – 18.

Mester, L., 1994b, Efficiency of banks in the third district. Federal Reserve Bank of Philadelphia, Working Paper, 94 – 1.

Mester, L. J. 1996. A study of bank efficiency taking into account risk preferences. Journal of Banking and Finance 20, 1025 – 1045.

Mishkin, F. S., 2004. The Economics of Money, Banking, and Financial Markets. Pearson publications, 7th ed.

Mitchell, K. and Onvural N. M., 1993. Fourier Flexible Cost Functions: An Exposition and Illustration Using North Carolina S&Ls. Working paper. North Carolina State University, June 1993.

Molyneux P., Altunbas Y., and Gardener E., 1996. Efficiency in European Banking, John Wiley and Sons.

Molyneux P., and Wilson, J., 2007. Developments in European banking. Journal of Banking and Finance, 31, 1907 – 1910.

Newbold, P. et al, 2007. Statistics for Business and Economics, 6th edition, Pearson.

Pasiouras, F (2008). Estimating the technical and scale efficiency of Greek commercial banks: The impact of credit risk, off-balance sheet activities, and international operations. Research in International Business and Finance 22, 301 – 318.

Pastor, J. M., 1999. Efficiency and risk management in Spanish banking firms: a method to decompose risk, Applied Financial Economics, 9, 4, 371 – 84.

Pastor, J. M., 2002. Credit risk and efficiency in European banking systems: A three stage analysis, Applied Financial Economics, 1 – 17, preview article.

Pastor, J.M. and Serrano, L. 2005. Efficiency, endogenous and exogenous credit risk in the banking systems of the Euro area. Applied Financial Economics, 15, 631 – 649.

Rao, A., 2006. Cost frontier efficiency and risk-return analysis in an emerging market. International Review of Financial Analysis 14, 283 – 303.

RBS Group 2006. Annual Report and Accounts.

Reifschnieder, D. and Stevenson, R., 1991. Systematic Departures from the Frontier: A Framework for the analysis of firm inefficiency. International Economic Review 32, 715 – 723.

Rhoads, S. A., 1998. The efficiency effects of Bank Mergers: An overview of case studies of Nine Mergers. Journal of Banking and Finance 22, 273 – 291.

Ritter, C., and Simar, L., 1997. Pitfalls of Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis, 8, 2, 167 – 182.

Rogers, K.E., 1998. Nontraditional activities and the efficiency of US commercial banks. Journal of Banking and Finance 22, 467 – 482.

Rose, P. S. and Hudgins S. C., 2008. Bank Management and Financial Services, 7th ed., McGraw Hill.

Rudin, W., 1976. Principles of mathematical analysis, Third edition. McGraw-Hill Inc., New York.

Sarafidis, V., 2002. An Assessment of Comparative Efficiency Measurement Techniques. Europe Economics, London.

Sathye, M., 2001. X-efficiency in Australian banking: An empirical investigation. Journal of Banking and Finance 25, 613 – 630.

Shcotter, A., 2008. Microeconomics: A Modern Approach. South-Western publication.

Shepard, R. W., 1970. Theory of Cost and Production Functions. Princeton University Press, Princeton, NJ.

Shaffer, S., 1988. A revenue-restricted cost study of 100 large banks. Federal Reserve Bank of New York, Research Paper 8806.

Shaffer, S., 1998. Functional forms and declining average costs. Journal of Financial Services 14, 2, 91 – 115.

Sheldon, G. 1999. Costs, competitiveness and the changing structure of European banking. Working paper, Foundation Banque de France pour la Recherche.

Shepherd, W. G. 1985. The Economics of Industrial Organization, 2nd edition, New York, Prentice-Hall.

Simar, L., and Wilson, P., 1998. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Management Science,44, 49 – 61.

Smith, P. 2008. Gauging the pressure. Accountancy Magazine.com, 26 – 29. Société Générale Group, 2007. Registration Document.

Spong, K., Sullivan, R., De Young., R. 1995. What makes a bank efficient? A look at financial characteristics and bank management and ownership structure. Federal Reserve Bank of Kansas City: Financial Industry Perspectives.

Stata Press Publication, 2005. Stata Longitudinal / Panel Data: Reference Manual Release 9. Stata Corp LP.

Stiroh, K. J., 2000. How did bank holding companies prosper in the 1990s? Journal of Banking and Finance 24, 1703 – 1745.

Streeter, B., 2008. Bad moves that don't fool anybody. ABA Banking Journal, March, 7 – 10.

Sunders, A. and Ingo, W., 1994. Universal Banking in the United States. Oxford University Press.

Taleb, N. 1997. The world according to Nassim Taleb, Derivatives Strategy 2, December/ January, 37 – 40.

The Banker, September 3rd 2007. Four pillars of support.

The Banker, January 2nd 2008. Facing up to the need of honesty.

The Economist, May 17th 2007. Risk and Reward: Worried about credit risk? You should fret more about pension funds than banks.

The Economist, Sept. 22nd 2007. Securitization: when it goes wrong. Tulkens, H., 1993. On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit. Journal of Productivity Analysis, 4, 183 – 210.

Valverde, S. C., and Fernandez, F. R., 2007. The determinants of bank margins in European banking. Journal of Banking and Finance 31, 2043 – 2063.

Valverde, S. C., et al, 2007. Do cross-country differences in bank efficiency support a policy of "national champions"? Journal of Banking and Finance 31, 2173 – 2188.

Vennet, R. V., 2002. Cost and Profit Efficiency of Financial Conglomerates and Universal Banks in Europe. Journal of Money, Credit, and Banking 34, 1, 254 – 282.

Weill, L 2004. Measuring Cost Efficiency in European Banking: A Comparison of Frontier Techniques. Journal of Productivity Analysis 21, 133 – 152.

Wheelock, C., and Wilson P. W., 2001. New evidence on returns to scale and product mix among US commercial banks. Journal of Monetary Economics 47, 3, 653 – 674.

White, H. 1980. Using Least Squares to Approximate Unknown Regression Functions. International Economic Review 21, 149 – 169.

Williams, J., 2004. Determining management behavior in European banking. Journal of Banking and Finance 28, 2427 – 2460.

Wooldridge, J. M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Wooldridge, J. M., 2003. Introductory Econometrics: A Modern Approach. Thomson South-western.

Yildirim, H, S., et al 2007. Efficiency of Banks: Recent evidence from the Transition Economies of Europe, 1993 – 2000. The European Journal of Finance 13, 2, 123 – 143.